THE UNIVERSITY *of* EDINBURGH

Biophysical Studies of Protein-ligand Interactions and the Discovery of FKBP12 Inhibitors

A Thesis Submitted for the Degree
of Doctor of Philosophy
by

Elizabeth A. Blackburn, B.Sc., M.Sc.

Institute of Structural and Molecular Biology
The University of Edinburgh
2009

Abstract

The principal aim of this study was to discover, through virtual screening, new non-immunosuppressive inhibitors for the human immunophilin FKBP12, a target of the immunosuppressant drugs rapamycin and FK506. The enzyme acts as peptidyl-prolyl isomerase catalysing protein folding in the cell. Structurally similar isomerase domains are important for molecular recognition in multi-domain chaperone proteins. FKBP inhibitors have been shown to have protective effects against nerve damage and are therefore interesting targets for the treatment of neurodegenerative diseases.

Virtual screening has been used to discover novel inhibitors for protein drug targets. Recent advances in computational power and the availability of large virtual libraries, such as the EDULISS database at Edinburgh University, have enhanced the appeal of this approach. X-ray structures of known protein-ligand complexes were examined to obtain an understanding of the key non-covalent interactions in the FKBP12 binding pocket. Virtual screening hits were selected using macromolecular docking and programs that employed a ligand-based approach. The bulk of the virtual screening in this study used Edinburgh University's in-house program LIDAEUS. In the course of this study nearly three hundred compounds were screened in the laboratory using biophysical and biochemical binding assays. Thirty four compounds were found to have an affinity for FKBP12 of less than one hundred micromolar.

To test virtual hits, it was necessary to select the most appropriate medium-throughput biophysical assay. The aim was to employ methods with sufficient sensitivity to detect compounds with affinity in the order of one hundred micromolar, coupled with the capacity to screen hundreds of compounds in a week. This study used a wide variety of biophysical techniques, these including: electrospray ionisation mass spectrometry, surface plasmon resonance and isothermal titration calorimetry. There was a particular emphasis on the quality of data from electrospray ionisation mass spectrometry. A correlation was found between the cone voltages that gave 50 % dissociation of the complex with the enthalpic contribution to the free energy of binding. From the careful examination of the differences in charge-state distributions between a pure protein and a protein-ligand mixture, it was possible to determine if a protein-ligand complex had been present in solution prior to dissociation during the electrospray process. This observation provides the basis for an assay that could be of general utility in detecting very weak inhibitors.

Declaration

The work presented in this thesis is the original work of the author, unless acknowledged to another source. This thesis has been composed by the author and has not been submitted in whole or part for any other degree.

Elizabeth A. Blackburn

Acknowledgements

I would like to thank my supervisor Prof. Malcolm Walkinshaw for his inspiration and guidance throughout this study. I would also like to thank all the members of the Walkinshaw group who have helped me, in particular Sandra Bruce, Steve Shave, Dr. Paul Taylor and Dr. Martin Wear.

I was fortunate to be sponsored by Organon (now part of the Schering-Plough Corporation) as part of a CASE studentship with the BBSRC. I would like to thank the members of the Computational Chemistry Group at Schering-Plough, Newhouse, in particular, Dr. John MacLean and Dr. Brad Sherborne for their enthusiasm and insightful advice. Thanks also to the Analytical Chemistry Group at Newhouse, in particular; Alistair Firth, Tim Liddicoat and Dr. Samantha Rutherford for allowing me so much freedom to experiment in their laboratories.

Thanks are also due to Dr. Kevin Bailey (Manchester University) for the synthesis of the KB2 series and to Dr. Victoria Butler (Glasgow University) for testing compound KB2_61D against *C.elegans*.

Finally, I would like to thank my family Alisan, Christina and Robin for their support and encouragement during my change of career. Robin has endured many late dinners and far too many whiteboard explanations. I would also like to thank Roley Walton for sparking my enthusiasm in Biology.

Abbreviations

2'-CMP	2’-Cytidine monophosphate
ADMET	Absorption, distribution, metabolism, excretion, toxicity
AMP	Adenosine monophosphate
ANS	Anilinonaphthalene-8-sulfonate
Bcl-2/-xl	B-cell lymphoma-2/xl
Bis-Tris	1,3-bis(tris(hydroxymethyl)methylaminopropane
BSA	Bovine serum albumin
C.elegans	*Caenorhabditis elegans*
ClogP	Calculated octanol water coefficient
CRM	Charge residue model
CS	Charge-state
CsA	Immunosuppressant drug, cyclosporin A
DEREK	Deductive Estimation of Risk from Existing Knowledge
DLS	Dynamic light scattering
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
DSS	Methylsulfinylmethylsulfanylmethane
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethyldiaminetetraacetic acid
EDULISS	Edinburgh University Ligand Selection System
ER	Endoplasmic reticulum
ESI-MS	Electrospray ionisation mass spectrometry
FK506	Immunosuppressant drug, also known as Tacrolimus
FKBP	FK506 binding protein
FKBP12	FK506 binding protein; 12 kDa isoform
FPLC	Fast protein liquid chromatography
HB	4-hydroxy-butanone
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
H-bond	Hydrogen bond
hcypA	Human Cyclophilin A; 18 kDa isoform
HEWL	Hen egg-white lysozyme
HIV	Human immunodeficiency virus
HPLC	High pressure liquid chromatography
Hsc70	Heat-shock chaperone 70 kDa
Hsp72/90	Heat-shock protein 72 or 90 kDa
HTS	High throughput screening
IC50	Half-maximal (50%) inhibitory concentration
IEM	Ion evaporation model
IP_3R	1,4,5-triphosphate receptor
IPTG	Isopropyl β-D-1-thiogalatopyranoside
ITC	Isothermal titration calorimetry

KB	Dr. Kevin Bailey (Manchester University)
K_d	Equilibrium dissociation constant
$K_{I,app}$	Apparent Equilibrium dissociation constant of an inhibitor
K_m	Michaelis constant
LIDAEUS	**Li**gand **D**iscovery **A**t **E**dinburgh **U**niversity
logP	Experimental octanol/water partition coefficient
MALDI	Matrix assisted laser desorption
MlogP	Moriguchi octanol/water partition coefficient
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
mTOR	Mammalian target of rapamycin
MTS	Medium throughput screening
MVSP	Multivariate statistical package (V3.13, Kovach Computing Services)
MW	Molecular weight
NAG3	N, N', N''-Triacetylglucosamine
NFAT	Nuclear factor of activated T-cells
nHBA	Number of hydrogen bond acceptor atoms in a molecule
nHBD	Number of hydrogen bond donor atoms in a molecule
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOS	Nitric oxide synthase
NPN	N-phenyl-naphthylamine
NTA	Ni^{2+}-nitrilo-triacetic acid
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG8000	Polyethylene glycol, mean molecular weight 8000 Da
PMSF	Phenylmethanesulfonylfluoride
PPIase	peptidyl-prolyl isomerase
rapa	Immunosuppressant drug rapamycin, also known as Sirolimus
Rheb	*Ras* homolog enriched in the brain
RMSD	Root mean squared deviation
RNaseA	Ribonuclease A
ROC	Receiver operating characteristic
Ro3	Rule of 3
Ro5	Rule of 5
RyR1	Ryanodine receptor, type 1
SAR	Structure activity relationship
SD1	San Diego compound 1, (2-[(4-methylphenyl)sulfanyl]-1-(morpholin-4-yl)ethan-1-one)
SDF	Structure-Data file
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SL	Steve Shave/Liz Blackburn series of compounds
SMILES	Simplified Molecular Input Line Entry Specification
SPR	Surface plasmon resonance
TDF	Thermal denaturation fluorescence
TGF-β	Transforming growth factor, β type

T_m	Mid-point protein melting temperature
TPR	Tetratricopeptide repeat
Tris	tris(hyroxymethyl)aminomethane
UFSRAT	Ultra fast shape recognition and atom typing
UPGMA	Unweighted pair-group method with arithmetic mean
VC50	Cone voltage that gives 50 % dissociation of a protein-ligand complex
Xaa	A naturally occurring amino acid

Contents

Figures

1 Chapter 1 Introduction

1.1 Scope of the project

This study focuses on discovering new small molecule inhibitors for the human immunophilin FKBP12 using virtual screening and biophysical techniques. High resolution x-ray structures of FKBP12-ligand complexes were used to predict key protein-ligand interactions and as templates for docking experiments. Hits from virtual screening programs were purchased and tested in the laboratory on recombinant protein using biophysical methods. The most active compounds were confirmed in a biochemical assay in which the inhibitor competed with a substrate for the active site of the protein. There was a particular emphasis on extending the sensitivity of biophysical methods to detect low affinity ligands of FKBP12.

1.2 Drug discovery

1.2.1 Exploring chemical space

Chemical space is vast; it has been estimated that there are 10^{60} drug-like compounds with up to 30 heavy atoms (Congreve *et al.*, 2008). All experimental screening methods rely on some rationale to provide a suitable number of compounds for testing in the laboratory. An experimental medicinal chemist might approach the problem from two directions. They may select a diverse range of compounds from a "real" compound library, or focus on the target protein by selecting or synthesising compounds chemically similar to known inhibitors/substrates (Snowden and Green, 2008; Diller, 2008). In the pharmaceutical industry the first stage in experimental compound testing often takes the form of a high-throughput screen (HTS) where many hundred thousand compounds are tested in a robotically assisted assay. As an alternative approach, virtual screening has the potential to sample a much larger portion of chemical space than any in-house "real" compound collection (Oprea and Matter, 2004). This is particularly true in a University setting where large compound libraries are not commonly held. Virtual screening aims to increase the probability of a test compound hitting the target protein when compared to a random selection; this process is known as enrichment (Halgren *et al.*, 2004; Verdonk *et al.*, 2004). In this study virtual screening took the place of HTS which was not possible for reasons of equipment and cost. It is of course possible to combine virtual methods in series with HTS to focus the screening library (Bajorath, 2002; Schnecke and Bostrom, 2006).

1.2.2 Structure-based drug discovery and virtual screening

Virtual screening methods are all based on some form of structural information. The observation that the arrangement of atoms of specific properties is important in drug design is not a new one. Alexander Crum-Brown wrote in 1869, "There can be no reasonable doubt that a relation exists between the physiological action of a substance and its chemical constitution, understanding by the latter the mutual relations of the atoms in the substance." (Crum Brown and Fraser, 1869). The difficulty lies in how completely we describe the geometric arrangement, the properties and the interactions between the atoms in the virtual model. Structure–based drug discovery using virtual screening may be split up into a series of stages; these are illustrated in Figure 1-1.

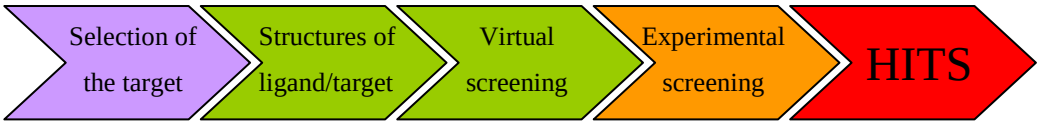


Figure 1-1 Structure-based drug discovery

In the first stage it is necessary to select a protein target or cellular receptor of therapeutic interest that can be produced in a stable form for experimental screening. Structures of the target and/or ligands are then assembled as input for the virtual screening program. The available structures influence the choice of virtual screening method. Virtual screening can be broadly divided into two classes; these are summarised in Figure 1-2 (together with the programs used in this study). The first class molecular docking uses a high-resolution structure of the target, these are usually x-ray or NMR structures of proteins or protein assemblies. Homology models have been used but are not considered optimum (Kitchen *et al.*, 2004). Structures of the target in complex with a ligand usually perform better than apo structures (Thomas *et al.*, 2006). In the second class, the ligand-similarity methods, it is not essential to know the target structure but structures of biologically active ligands are necessary (Willett, 2006; Eckert and Bajorath, 2007).

There is inevitably a trade off between complexity of the model and the speed of computational prediction. The preference in this project was to use high-throughput (HT) virtual screening programs. HT programs aim to encode the key features of a molecule and inter-molecular interactions to make fast but useful predictions; choice of these key parameters is not a trivial problem. The goal of HTS programs is to screen from thousands to millions of compounds in days.

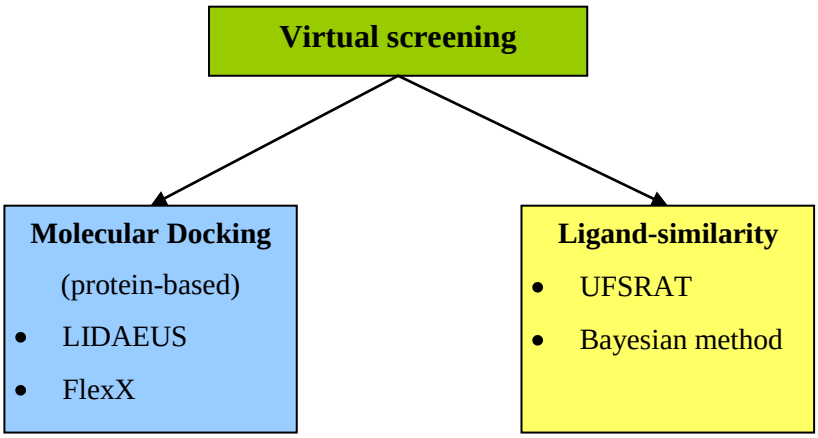


Figure 1-2 Virtual screening methods used in this project

Four different programs were used for virtual screening in this project, two protein-based and two ligand-based. Docking experiments were performed using the programs LIDAEUS (Wu *et al.*, 2003) and FlexX (Rarey *et al.*, 1996). Here the aim of experiments was to match a 3D description of the test ligand to a 3D description of the target protein, score the potential interaction and rank ligands by score. UFSRAT, a ligand similarity based method, was used to compare the shape and polarity of known inhibitor of FKBP12 to molecules in a virtual library (Steven Shave, Edinburgh University). This method was not dependent on knowing the structure of the target protein. Finally, a Bayesian machine learning method was used to assign a relative activity probability to a compound in a library (Angelopoulos *et al.*, 2009). Pre-calculated physicochemical/structural descriptors and experimental activity data from a set of known actives and inactives of FKBP12 were used to train the algorithm. Like UFSRAT, the Baysian method did not rely on knowledge of the structure of the protein target. The method can be distinguished from UFSRAT in that the model includes experimental affinity data for many rather than one known inhibitor of FKBP12.

In the final stage, ligands selected by a virtual screening program are purchased and tested in an experimental assay. This project used a wide variety of biophysical techniques, these included: electrospray ionisation mass spectrometry, surface plasmon resonance and isothermal titration calorimetry. Biochemical and biophysical assays are discussed in Chapters 3, 4, 5 and 6. The experimental results for compound collections are discussed in Chapter 7.

1.3 Virtual screening databases – EDULISS 2

EDULISS 2 (**ED**inburgh **U**niversity **L**igand **S**election **S**ystem) was the principle database of virtual small molecules used in this study. EDULISS (Hinton, 2005; Hsin, 2009) is a relational database comprising over 6 million compounds from commercial and academic compound collections (http://eduliss.bch.ed.ac.uk/eduliss/, Jan 2009). The aim of the database was to store 3D structures for virtual screening experiments with a very wide variety of associated molecular descriptors that are thought likely to correlate with biological activity. The energy minimised 3D coordinates for each molecule were calculated from the 2D coordinates provided by the compound supplier using the program CONCORD (TRIPOS). 1665 topological, geometric, physicochemical and toxicological descriptors were calculated from 2D structure for each molecule in the database using the programs DRAGON 5.4 (Todeschini R. and Consonni V., 2005) and DEREK (Sanderson and Earnshaw, 1991). Of the 6 million compounds, some 4.3 million fit the Lipinski "Rule of 5s" criteria (Lipinski *et al.*, 1997); 3.2 million the Oprea lead-like criteria (Hann and Oprea, 2004) and 230,000 the Astex "Rule of 3" (Carr *et al.*, 2005). Drug-like criteria are discussed in 1.4.

A novel method for identifying unique compounds in the database was developed by Kun-Yi Hsin (PhD Thesis, Edinburgh University, 2009). This method relies on distinguishing compounds using a small number of particularly discriminatory descriptors. These include a 3D-Wiener index, an electronegativity descriptor and a polarisability descriptor. Of the 6 million compounds in the database, 3.8 million are unique.

Hsin has developed a web-based interface for EDULISS 2 that provides a convenient way of extracting families of compounds with a user-defined set of properties. It was straightforward to apply a filter for drug-likeness when selecting compounds from EDULISS 2. The virtual compound sub-library was extracted in the form of a concatenated SDF file that was then piped into the virtual screening program. It was possible to append the SDF file to include a field for any descriptor of interest.

1.4 Screening compounds for drug-likeness

Towards the end of the last century it became apparent that many drugs were failing at a relatively late stage in the drug development pipeline due to unfavourable ADMET properties (absorption, distribution, metabolism, excretion and toxicity). This is obviously inefficient in terms of time and money. Much effort has gone into predicting ADMET

properties using *in-silico* methods (Norinder and Bergstrom, 2006). Experimentally measured physicochemical properties and toxicity data are not available for many compounds in HTS screens and virtual databases. A simple solution to the ADMET problem is to use a selection of these physicochemical parameters to determine whether a molecule is "drug-like" or not.

Christopher Lipinski undertook survey of physicochemical properties of known orally delivered drugs (Lipinski *et al.*, 1997). He defined "drug-like" molecules as those that met 3 out of 4 of the following rules: a molecular weight ≤ 500 Da, ≤ 5 hydrogen bond donors, ≤ 10 hydrogen bond acceptors, the sum of N's and O's ≤10 and a ClogP ≤ 5 (MlogP ≤ 4.15). These criteria are known as Lipinski's "Rule of 5s" (Ro5), poor absorption and permeation are more likely if they are not met. ClogP is a calculated log of the octanol/water partition coefficient (P) from structure. MlogP is a commercial algorithm performing this function (Moriguchi I. *et al.*, 1992). LogP reflects the relative hydrophobicity of the molecule. Hydrophobicity is a factor in the oral bioavailability and the partitioning of a drug within different cellular compartments. More stringent sets of criteria have been proposed for initial stages of drug discovery. For example; Lead-likeness restricts MW < 350 Da and ClogP < 3 (Teague *et al.*, 1999).

Astex "Rule of 3" (Ro3) is a set of criteria for describing fragment like molecules: MW ≤ 300 Da, total polar surface area ≤ 60 $Å^2$ and ClogP ≤ 3 (Carr *et al.*, 2005). The rationale for choosing lower parameter values for lead-like and fragment-like compounds is that compound optimisation almost inevitably increases the molecular weight, complexity and hydrophobicity of the molecule by the addition of functionality (Oprea *et al.*, 2007). It is therefore important to start drug development with molecules that are small, simple and not too hydrophobic.

Experimental data for the above criteria were not available for the majority of the compounds in EDULISS 2. This meant it was necessary to base any decision on compound inclusion on a descriptor calculated from structure. An interesting observation is that the best known and most potent inhibitor of FKBP12 fails Lipinski's criteria for more than one descriptor. Rapamycin has a molecular weight of 914 Da and 13 hydrogen bond acceptors; XlogP = 4.3 and MlogP = 1.98 (Cheng *et al.*, 2007). Calculated values of logP (ClogP) are generally atom/group additive with the inclusion of a correction factor. ClogP can vary considerably for a specific compound when computed using different algorithms. Moreover,

FKBP12 has a hydrophobic binding site and one could predict that effective inhibitors would be relatively hydrophobic molecules (+ logP). Drugs derived from natural products fail Lipinski's criteria more frequently than synthetic small molecules (Ganesan, 2008). However, perhaps the primary consideration was whether the compound would be soluble enough in aqueous buffer to use in a laboratory assay. This is discussed further in Chapter 7.

1.5 Docking

Virtual screening has had some success in discovering novel inhibitors for protein drug targets. Taylor *et al.* (2008) include a table of nine studies where small molecule inhibitors (low micromolar to sub-nanomolar affinity range) were discovered using the programs: Catalyst (Greene *et al.*, 1994), DOCK (Ewing *et al.*, 2001), FlexX (Rarey *et al.*, 1996), GOLD (Jones *et al.*, 1995) and LIDAEUS (Wu *et al.*, 2003). Of the programs listed above, DOCK, FlexX, GOLD and LIDAEUS are docking programs (Catalyst is a pharmacophore modelling program). Two docking programs were used in this study LIDAEUS and FlexX. This section attempts to summarise (very briefly) a few key features of such programs.

Kuntz *et al.* developed the first docking programs in the early 1980s (Kuntz *et al.*, 1982). Since then, there has been a divergence of approach to tackling the docking of small molecules into proteins. However, the workflow for most programs can be divided up into four basic stages. These are illustrated in Figure 1-3.

In the first stage, a 3D structure of the target protein is inputted; generally in the form of a PDB file, giving 3D coordinates and connectivity information from the x-ray structure of a protein in complex with a ligand. The active site of the protein is defined and a description of the site generated, this description can be atomic, surface or grid (Brooijmans and Kuntz, 2003). In the second stage, the putative ligand is matched or docked into the active site; the orientation of the ligand in the active site is referred to as a "pose". In the third stage, the pose is scored. The docking of the ligand into the active site is guided by the posing/scoring algorithm. The algorithm accepts or rejects a change in ligand pose based on a score. In the final stage, an estimate is made of the free energy change of complex formation; this is used to rank the ligands.

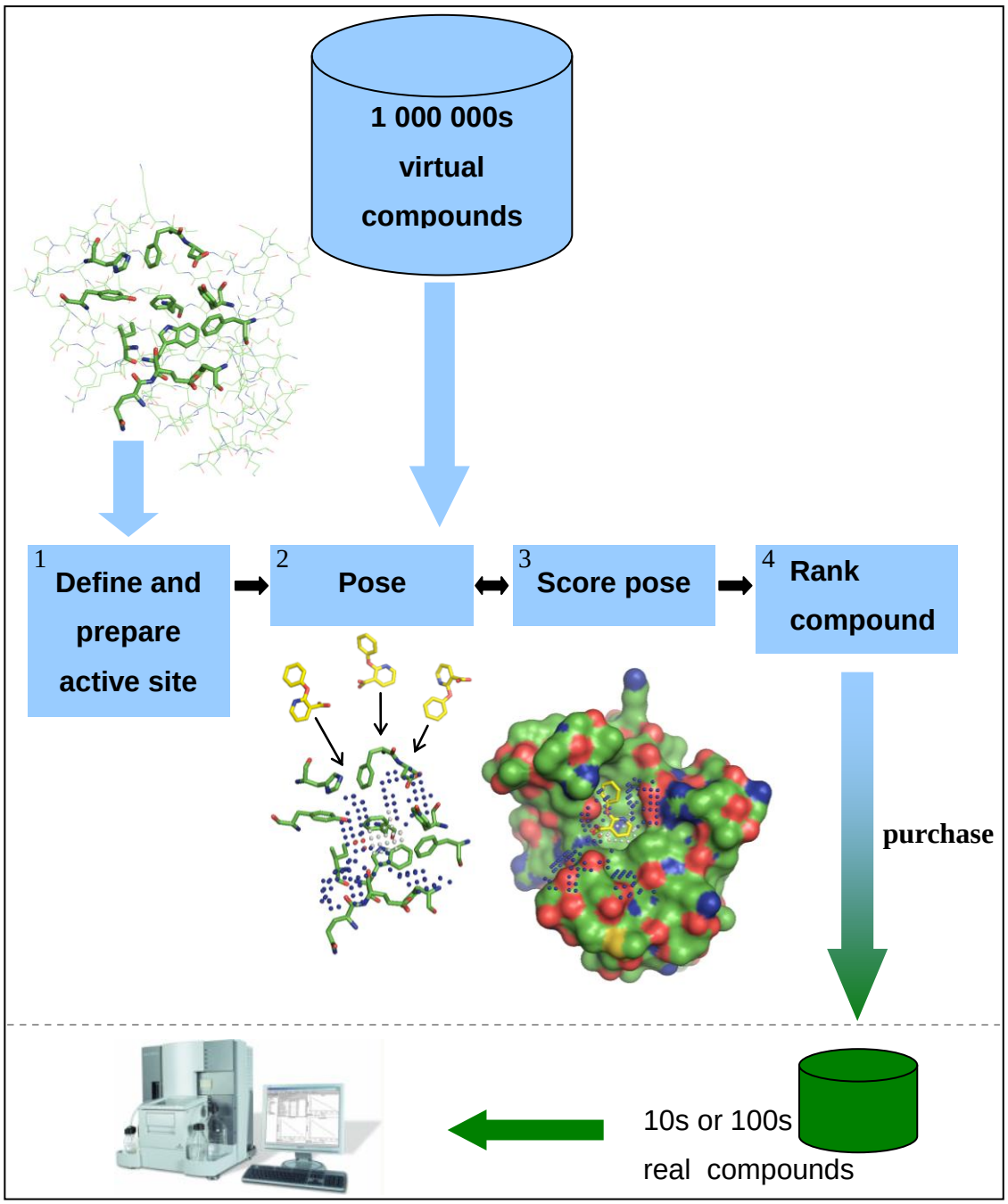


Figure 1-3 The four stages in a docking experiment

To make the process of docking computationally efficient, it is necessary to limit the possible number of orientations of the ligand in the active site and use a strategy that efficiently samples the space. When docking small molecules, this is most often done by describing the site as a grid of "site-points". Each site-point is assigned a potential energy value calculated from the protein receptor at that point (Goodford, 1985). During the docking process, the site-points are matched to atoms in the representation of the ligand using some form4 of systematic search.

Much consideration has been given to predicting the correct conformation of the ligand in the active site. Many programs tackle this problem by breaking the ligand up into fragments. A base fragment of relatively rigid conformation is selected and docked in the best scored location; the ligand is then reconstructed fragment by fragment (Rarey *et al.*, 1996; Jones *et al.*, 1997). This approach is used by DOCK, FlexX and Glide (Friesner *et al.*, 2004). Alternatively, small random changes can be made to the ligand using Monte Carlo techniques. GOLD uses a genetic algorithm to achieve a solution for the most energetically favourable conformation of the ligand in the active site. LIDAEUS tackles the problem by using a fuzzy definition of "fit" for a minimum energy conformation of the ligand to site-points.

Scoring functions

Many scoring algorithms in fast docking programs such as LIDAEUS use empirical methods, similar to those described by Boehm (Klebe and Boehm, 1997; Boehm, 1998). Binding energies are calculated by summation of individual terms for different contributions to binding energy such as; ionic interactions, hydrogen bonds and van der Waals forces. In Boehm's work empirical weighting constants were derived from crystallographic data.

1.6 Tackling the scoring problem

At the time of writing, it is still not possible to predict the absolute free energy of ligand binding with any degree of precision using the scoring functions of high-throughput virtual screening programs (Perola *et al.*, 2004; Feher, 2006; Konstantinou-Kirtay *et al.*, 2007; Waszkowycz, 2008). This is in part due to the challenges of fully predicting entropic factors and to the extremely complex nature of the interaction energy landscape (Klebe and Boehm, 1997; Friesner *et al.*, 2006). However, docking programs do hope to rank a series of compounds so as to increase the probability of activity in the highest scoring hits. This inevitably means that when virtual hits are tested in the laboratory a significant portion of the hits are found to be false positives or of relatively low affinity for the target. Drug discovery is an iterative process, in which virtual screens do not discover final products but aim to provide new starting points for lead development. It is therefore advantageous to couple any virtual screening program to sensitive experimental assays to extract useful structure-activity relationships from low affinity ligands. This has become increasingly important with the recent practice of selecting smaller lead compounds. Less chemically complex lead compounds allow the medicinal chemist the most scope to optimise ADMET properties early

in the drug pipeline. However, smaller leads or fragment-like compounds have a reduced potential for making interactions with the target.

This study tackled the scoring problem from two directions. Firstly, by careful consideration of "hot spots" in the active site of the protein during virtual screening and secondly, by increasing the throughput and optimising the sensitivity of biophysical measurements:

- A careful analysis was made of non-covalent interactions in known FKBP12 inhibitor complexes to discover amino acids in the active site that made conserved interactions with ligands (described in section 1.11.5). This information was used at two stages in the docking process. Firstly to guide the site point generation in LIDAEUS and secondly to select compounds from the final selection of high scoring docked poses that made the most conserved non-covalent interactions.

- Biophysical techniques assays were developed for testing inhibitors of FKBP12 on a 96 well plate format. There was particular emphasis placed on the evaluation of electrospray ionisation mass spectrometry (ESI-MS) as a sensitive tool to detect weak non-covalent interactions.

1.7 Non-covalent interactions

That non-covalent interaction is pivotal to biomolecular structure, molecular recognition and drug-design is not disputed. However, a full description and deconvolution into individual interactions, even in the context of a high resolution crystal structures, is complex. The scoring functions of high-throughput docking programs aim to give a fast prediction of the energy of the interaction between a posed ligand and a protein. It is therefore useful to discuss the nature of non-covalent interactions. Non-covalent interactions are an order of magnitude or two weaker than standard covalent bonds and can exist over much greater distances. It is the relatively low energy of a binding interface (20 kcal·mol^{-1} to 0.5 kcal·mol^{-1}) and the relative ease of making and breaking, that makes such bonding so important to biological systems (Boehm, 1998; Cerny and Hobza, 2007).

It would be convenient to place non-covalent interactions into groups based on the physical origin of the attractive or repulsive forces that govern the interaction. A very broad division might place intermolecular forces into two categories: purely electrostatic arising from a Coulombic attraction between full or partial charges, and polarisation forces dependent on an

induced dipole (permanent or rapidly fluctuating). Such a distinction can be useful when deconvoluting bond energies in a vacuum, but in an aqueous environment the polar nature of water and its capacity to form networks of hydrogen bonds, both in the active site and around a solvated ligand, needs to be considered. Non-covalent interactions are often divided into electrostatic, hydrogen bonds, hydrophobic, attractive van der Waals dispersion forces and van der Waals repulsive forces. Table 1-1 summarises bond lengths and energies for this broad classification.

Care must be taken with terming the hydrophobic effect a non-covalent interaction as it is really a macroscopic effect. It results from differences in hydrogen bonding networks in the environment of the protein's active site and the bulk solvent, considered with the associated entropic effects (Sturtevant, 1977; Spolar *et al.*, 1989). It is also difficult to assign typical bond energy, for example, hydrogen bonds can have a partial covalent character. Hydrogen bonds have much less wave function overlap than is typical for a covalent bond that generally has a bond distance less than 2 Å (Isaacs *et al.*, 1999). In general, the shorter and the closer to 180^{o} the hydrogen bond, the more covalent in character and the larger the bond energy.

Type of non-covalent interaction	***Specific example***	***Equation for energy of interaction***	***Typical Energy (kcal·mol^{-1})***	Typical distance r_{ij} (Å)	***Ref***
Electrostatic	monopole:monopole -COO^{-}····$^{+}$H$_3$N-	$\frac{q_i q_j}{\varepsilon r_{ij}}$	-7	2.8	(Price *et al.*, 2007)
Hydrogen bond (neutral)	>N-H····O═C<	$\frac{q_i q_j}{\varepsilon r_{ij}} + \frac{C}{r_{ij}^{12}} + \frac{D}{r_{ij}^{10}}$	-0.4 to – 1.1	3.0 (2.5-3.2)	(Boehm, 1998)
Dispersion van der Waals (attractive)	≻C-H····H-C≺	$\frac{A}{r_{ij}^{6}}$	-0.12/A^2	3.5 (3.2-3.8)	(Burkhard *et al.*, 2000)
Steric repulsion van der Waals	≻C-H····H-C≺	$-\frac{B}{r_{ij}^{12}}$	+, very high	less than the sum of the van der Waals radii	(Bondi, 1964)

Table 1-1 Non-covalent interactions

A summary of energies of interaction and bond distances for a selection of non-covalent interactions. ε is the dielectric constant of the medium, r_{ij} represents the atom distance between atoms i and j, q_i and q_j their respective charges and A, B, C and D constants.

When examining an x-ray structure to determine the interactions between a ligand and protein, the usual approach is to measure the distance between a pair of heavy atoms (atoms other than hydrogen in this context) involved in the putative interaction, X-H...X. Where: X represents a heavy atom that can act as a hydrogen bond donor or acceptor; these atoms are O, N and less frequently C. If the distance is between 3.2 and 2.5 Å and the appropriate atoms are involved a hydrogen bond may be assigned (Boehm, 1998). This is not necessarily clear cut, for example, a distance of up to 3.6 Å for C-H...O can be considered an acceptable criterion for a hydrogen bond if the angle made by the three atoms is close to 180^{o}. For electrostatic interaction formal charges on each atom in the pair are present and the distance is shorter than that for a hydrogen bond. For atom pairs where the distance is less than 3.8 Å and a hydrogen bond is not an appropriate conclusion, a van der Waals interaction is considered.

1.8 Thermodynamic considerations in drug discovery

The equilibrium of a protein P with ligand L to form complex PL can be described by the equation:

$$P + L \rightleftharpoons PL$$

It is usual in biochemistry to quantify the affinity of the ligand for the protein by the equilibrium dissociation constant (K_d). Equation 1 describes the ratio of the species at equilibrium. This relationship is known as the law of mass action.

$$K_d = \frac{[P]\bullet[L]}{[PL]} \qquad \textbf{Equation 1}$$

For a ligand to bind to a protein under a given set of conditions there must be a negative Gibbs free energy change on ligand binding ($\Delta G_{binding}$). $\Delta G_{binding}$ is made up of both enthalpic (ΔH) and entropic terms (ΔS), equation 2

$$\Delta G_{binding} = \Delta H - T\Delta S \qquad \textbf{Equation 2}$$

The Gibbs free energy change on binding is related to the equilibrium dissociation constant. This is known as the van't Hoff relationship (Atkins, 1986).

$$\Delta H - T\Delta S = RT \ln K_d \qquad \textbf{Equation 3}$$

Where: R the universal gas constant and T the absolute temperature in Kelvin.

The term which dominates the free energy change on binding is largely due to the polar nature of the participating chemical groups and their conformational freedom. Enthalpic contributions to bonding are principally dependent on electrostatic attractive interactions such as hydrogen bonds and van der Waals forces. However, there is an unfavourable enthalpy associated with breaking hydrogen bonds between water molecules and the polar groups of the ligand when the ligand is removed from the solvent (desolvation). 8 $kcal \cdot mol^{-1}$ is a recent estimate of the enthalpy of desolvation associated with a polar group (Freire, 2008). Hydrogen bond energies between ligand and protein are very sensitive to bond geometry and distance. Van der Waals interactions are particularly dependent on the geometric fit of the ligand in the active site (Cooper, 2005). Therefore the ligand must make "good" hydrogen bonds/electrostatic interactions to overcome desolvation effects if there is to be a negative free energy change when the ligand binds. It might be argued that a very favourable enthalpic term can deliver specificity for a target due to the influence of the topography of the active site.

The entropy of binding can be broadly divided up into desolvation entropy and conformational entropy (Sturtevant, 1977). It is important to consider both the ligand and the protein in the above terms. Solvated unbound ligands have a shell of ordered waters with stronger hydrogen bonding than those found in the bulk solvent. Desolvation produces a favourable entropic term as the ordered water molecules in the solvation shell lose a hydrogen bonding network and become less ordered. From the perspective of the protein, there can be a favourable entropic term from the displacement of waters in the active site. Crystal structures frequently show conserved hydrogen bonded waters in the active site that are displaced on ligand binding. The apo structure of FKBP12 shows three ordered waters in the active site, labelled W1, W2 and W3 in Figure 1-9. In addition, proteins are not rigid structures and binding of a ligand to a protein may "tighten the structure" and lead to a loss in translational and rotational entropy. There are also similar losses of entropy for the ligand on binding (Chang *et al.*, 2007).

In the context of drug discovery, one must always consider the strength of a new hydrogen bond in the lead optimisation process relative to the enthalpic desolvation of the ligand (Cabani *et al.*, 1981). An additional electrostatic interaction may also reduce the conformation entropy of the complex. It can be argued that targeting new hydrogen bonds to conformationally flexible regions of the protein would confer the maximum unfavourable conformational entropy changes. An enthalpic gain can often be accompanied by an entropic

loss, meaning little change in ΔG. An interesting example is described by Freire's group, where, in the optimisation of an HIV-1 protease inhibitor, a thioether group was replaced by a sulfonyl group. This resulted in an additional strong hydrogen bond to an aspartic acid residue. The binding enthalpy was improved by 3.9 $kcal \cdot mol^{-1}$, however this was balanced by an entropy loss that resulted in no increase in affinity (Lafont *et al.*, 2007). This is an example of what is often described in the literature as enthalpy/entropy compensation. This term is also used to describe the increase in ΔH and the concomitant decrease in a favourable entropic contribution (ΔS) that is generally observed as temperature is increased due to changes in heat capacity on ligand binding at constant pressure (ΔC_p) (Cooper *et al.*, 2007). Medicinal chemists can try to restrict an unfavourable entropic change on ligand binding by designing conformationally restricted ligands. It has been suggested that the macrocyclic immunosuppressive ligands of the immunophilins; rapamycin, FK506 and cyclosporin, have increased affinity for their target due to the restricted conformational flexibility afforded by the ring (Driggers *et al.*, 2008).

1.9 FKBP12 as a model protein for structure based studies

FKBP12 was selected as a suitable target for a project combining virtual and experimental methods. The 12 kDa, 108 amino acid protein is present in the cytoplasm and is the most abundant FKBP in humans. FKBP12 can be considered a model protein in the FKBP family of immunophilin proteins. It represents the smallest member of the family and at least one FKBP12-like domain is found in all other family members. There are 15 human FKBPs currently identified; they possess a range of functionally significant domains that flank the PPIase domain, and are distributed across almost all tissues (Rulten *et al.*, 2006; Somarelli *et al.*, 2008). FKBP12 domains are considered a therapeutic target for a number of neurodegenerative conditions (Valentine *et al.*, 2007). There are particularly high levels of FKBP expression in the brain and increased expression in areas of neuronal damage (Steiner *et al.*, 1992; Achim *et al.*, 2004). FKBP inhibitors have been shown to have protective effects against nerve damage and in some animal models to aid neuronal re-growth (Poulter *et al.*, 2004). The non-immunosuppressive ligand GPI1046 has been shown to have a neuroprotective effect in an animal model of nerve damage in HIV induced dementia (Steiner *et al.*, 2007). High-resolution X-ray data is available in the Protein Data Bank for FKBP12 in complex with a variety of small molecules and there is a wide body of literature (Berman *et al.*, 2000; Dornan *et al.*, 2003). There are well established protocols for protein production in *E.coli* developed at Edinburgh University (Wear *et al.*, 2007b).

1.10 FKBP- peptidyl-prolyl isomerase, immunophilin or chaperone?

FKBP12 has been studied from two angles; from a biological standpoint as a peptidyl-prolyl isomerase (PPIase) and from a pharmacological perspective in its role in the immunosuppressive action of the drugs FK506 and rapamycin (Harding *et al.*, 1989; Braun *et al.*, 1995). The term peptidyl-prolyl isomerase and immunophilin are used somewhat interchangeably to describe the FKBPs. The immunophilin family are additionally described as chaperone proteins. This is in part because the biological roles for the FKBPs have not been fully established. All FKBPs contain at least one FKBP12-like PPIase domain, with varying degree of affinity for FK506. It is useful to start by considering the repertoire of proteins with PPIase sequences.

The *cis-trans* isomerisation of the peptide bond preceding proline has been shown to be a rate-limiting step in protein folding within the cell (Wedemeyer *et al.*, 2002). Proteins that catalyse the conversion of this bond between the *cis* and *trans* forms are known as peptidyl-prolyl isomerases (PPIases). PPIases are ubiquitous proteins that can be present in concentrations of up to 1 % in the cytosol (Hacker and Fischer, 1993). They form a highly conserved multigene family found in a wide variety of prokaryotes and eukaryotes. PPIases are divided into three structurally distinct groups: cyclophilins, FKBPs and parvulins (Galat, 2003). Interestingly, each family of proteins possess a unique proline binding site, a lack of sequence homology and have not been found to have common ligands beyond small solvent molecules (Burkhard *et al.*, 2000).

The term immunophilin is used to describe the cyclophilins and FKBPs. The immunophilins were pharmacologically characterised by their ability to bind a range of immunosuppressant drugs that take the place of proline in the active site. Cyclophilins bind cyclosporin A and the FKBPs, rapamycin and FK506 (amongst other structurally related macrocyclic lactones). The earliest immunophilins to be identified were the small single domain proteins; human cyclophilin A (18 kDa) and FKBP12 (12 kDa) (Handschumacher *et al.*, 1984; Harding *et al.*, 1989; Siekierka *et al.*, 1989). Most eukaryotes have a homologue to these proteins (Page *et al.*, 1995). It has been shown that the immunosuppressant drugs do not exert their effect by inhibiting the PPIase action of the immunophilins, but by acting as chemical inducers of protein dimerisation, bringing together two proteins and forming an immunophilin-drug-protein ternary complex. It is the inhibition of the second protein that modulates a specific cellular signalling pathway leading to the inhibition of T-cell activation (Hamilton and

Steiner, 1998). In this study the aim was to find ligands that bound to the PPIase domain of FKBP12 and do not form an immunosuppressive complex.

Even considering the simplest of organisms there are a significant number of isoforms within both the cyclophilin and FKBP12 families; *C.elegans* has 18 cyclophilins and 8 FKBPs (Bell *et al.*, 2006). The larger immunophilins are multi-domain proteins and their binding partners have not been fully characterised. The biological role of an FKBP is often mediated through an interaction with a domain distinct from a PPIase domain. The variety of human FKBPs is discussed in more detail in the next section. These observations would indicate a high degree of redundancy in PPIase activity and lead one to ask the question: why are there so many different proteins performing the same role in the cell? These observations suggest that FKBPs have evolved diverse biological roles (Somarelli *et al.*, 2008). Current evidence would suggest that the defining feature of the immunophilins is not PPIase activity but their role as chaperones. It is their ability to recognise a specific structural motif or mis-fold in a protein partner that unites them (Ivery, 2000). The biological roles for the immunophilins have not been completely elucidated, however, they do include the ability to catalyse the refolding of partially denatured proteins, the stabilisation of multiprotein complexes and chaperone activity (Barik, 2006). Following these observations one could argue that the immunophilin ligands cyclosporin A, FK506 and rapamycin act as peptidomimetics. Microbial metabolite antibiotics such as FK506 are perhaps evolved to target proteins that are important to the housekeeping functions of the host species (Driggers *et al.*, 2008).

Although inhibition of PPIase activity may not be central to the mode of action of novel FKBP12 inhibitors; structural insights into the mechanism of PPIase action and the dissection of non-covalent interactions with existing ligands can only enhance the study of potential immunophilin binding partners. More practically, measuring the PPIase activity of a sample of recombinant protein is a good way of assessing if it behaves in a similar manner to the native protein. Likewise, monitoring the inhibition of isomerisation of a known substrate is helpful in determining the specificity of an inhibitor for the active site.

1.11 The FK506 binding proteins

1.11.1 Biological diversity and distribution

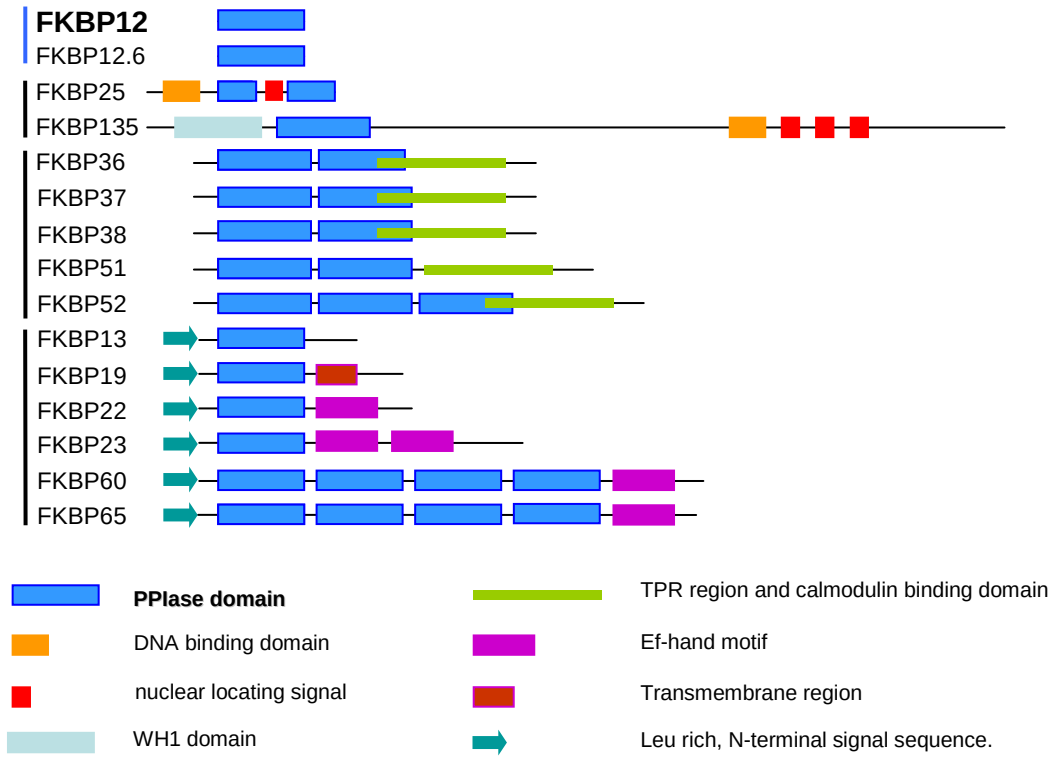


Figure 1-4 Domain structure of Human FKBPs.

FKBPs are characterised by the presence of at least 1 FKBP12-like PPIase domain. They are usually subdivided by considering the functionally significant domains flanking the PPIase domain and their cellular location (Himukai *et al.*, 1999a). The schematic representations in this figure use the molecular weight naming system; the black vertical bars indicate the FKBP subdivision.

FKBPs are found in all kingdoms (www.uniprot.org). The total number identified in a given organism is dependent on the method of classification. Bioinformatic categorisation relies on sequence homology to the FK506 binding domain of FKBP12 (PPIase-like domain). Not all FKBPs have been structurally characterised. For those that have, the majority show PPIase activity. Surveys by Rulten and Somarelli identify 15 FKBPs in humans (Rulten *et al.*, 2006; Somarelli *et al.*, 2008). Galat lists 17 human FKBPs; this author places a greater reliance on sequence homology and less on identification of mRNA sequences or protein expression (Galat, 2008). The earliest naming convention used FKBP followed by the approximate molecular weight in kDa. The naming system for FKBP genes uses *FKBP* (sometimes *FKB*) followed by a number (and in some cases a letter); FKBP12 is the protein of the gene *FKBP1A*. When the gene name is not denoted using italics, the two naming systems can be confusing.

FKBP ***(gene)***	**Family**	**Location**	**Binding partners**	**X-ray /NMR structure**	**References**
FKBP12 *(FKBP1A)*	cytoplasmic	ubiquitous, brain	RyR1, TGF-β	2pnn.pdb$_{x\text{-}ray}$,	(Harding *et al.*, 1989)
FKBP12.6 *(FKBP1B)*	cytoplasmic	ubiquitous, brain, thymus	RyR2	1c9h.pdb$_{x\text{-}ray}$	(Deivanayagam *et al.*, 2000)
FKBP25 *(FKBP3)*	nuclear	nucleus	histone deacetylase 1 and 2 (N-term region of FKBP25)	1pbk.pdb$_{x\text{-}ray}$	(Galat *et al.*, 1992; Liang *et al.*, 1996; Yang *et al.*, 2001)
FKBP135	nuclear	*nucleus*	actin filament assembly?	no struc.	(Ishikawa *et al.*, 1998)
FKBP36 *(FKBP6)*	TPR-domain	widely expressed, testes	clathrin, Hsp72, GADPH, Hsp90, Hsc70; no PPIase activity	3b7x.pdb (FKBP12-like domain)	(Jarczowski *et al.*, 2008a; Jarczowski *et al.*, 2008b)
FKBP37 *(AIP)*	TPR-domain	*cytoplasmic, liver*	putative?	no struc.	(Kuzhandaivelu *et al.*, 1996)
FKBP38 *(FKBP8)*	TPR-domain	ubiquitous, mitochondrial membrane; brain	calcineurin (without FK506), Bcl-2,Bcl-xl, Herg, no PPIase activity	2awg.pdb$_{x\text{-}ray}$	(Lam *et al.*, 1995; Wang *et al.*, 2005; Bai *et al.*, 2007)
FKBP51 *(FKBP5)*	TPR-domain	cytoplasm, nuc.	progesterone rec. complex with Hsp90	1kt1.pdb$_{x\text{-}ray}$	(Nair *et al.*, 1997; Sinars *et al.*, 2003; Banerjee *et al.*, 2008)
FKBP52 *(FKBP4)*	TPR-domain	cytoplasm, nuc.	progesterone rec. complex with Hsp90 and p53; glutocorticoid receptor	1q1c.pdb$_{x\text{-}ray}$ (N-term: 1-260)	(Peattie *et al.*, 1992; Wu *et al.*, 2004; Davies and Sanchez, 2005; Chadli *et al.*, 2008)
FKBP13 *(FKBP2)*	ER	ER lumen	C-chain complement C1q, 4.1G (membrane protein)	2pbc.pdb$_{x\text{-}ray}$	(Nigam *et al.*, 1993; Walensky *et al.*, 1998; Neye and Verspohl, 2004)
FKBP19 *(FKBP11)*	ER	*ER lumen, pancreas*	putative single-pass membrane domain?	no struc.	(Rulten *et al.*, 2006)
FKBP22 *(FKBP14)*	ER (EF-hand domain)	ER lumen, heart	putative?	no struc.	(Patterson *et al.*, 2002b)
FKBP23 *(FKBP7)*	ER (EF-hand domain)	*ER lumen, heart*	putative?	no struc.	(Nakamura *et al.*, 1998)
FKBP60 *(FKBP9)*	ER (EF-hand domain)	*ER*	putative?	no struc.	(Rulten *et al.*, 2006)
FKBP65 *(FKBP10)*	ER (EF-hand domain)	ER lumen	tropoelastin and collagen (growth and repair?)	no struc.	(Patterson *et al.*, 2002a; Patterson *et al.*, 2005)

Table 1-2 Human FKBPs

Information collated from searches in PubMed (www.ncbi.nlm.nin.org), UniProt (www.uniprot.org) and the PDB (Berman *et al.*, 2000); December 2008. "*Location" of FKBP in italics indicates that the cellular location was determined from mRNA expression sequences.*

FKBPs are present in almost all human tissues often with high levels of expression. Mammalian FKBPs are generally subdivided into 4 groups; cytoplasmic, nuclear, TPR-domain (tetratricopeptide repeat) and ER (Himukai *et al.*, 1999b). This classification is based on functionally significant regions, flanking or contained within the PPIase domains. Figure 1-4 shows the domain structure of the 15 human FKBPs. Table 1-2 provides a summary of human FKBPs, cellular localisation, binding partners and details of x-ray/NMR structures. This is large area of research and the following paragraphs only provide a very brief summary of research to date. Figure 1-5 shows an alignment of FKP12-like PPIase domains from 8 human FKBPs deposited in the PDB (Berman *et al.*, 2000). The PPIase domain of FKBP12 is discussed in more detail in section 1.11.5.

It can be seen from Table 1-2 that the cytoplasmic and the TPR domain FKBPs are the best characterised. The TPR domain FKBPs, excluding the poorly characterised FKBP37, all form a component of a multi protein complex. Hsp90 (heat shock protein 90) has been shown to bind to a TPR domain of FKBP51, 52 and 36. FKBP51 and FKBP52 form complexes with the progesterone receptor and a variety of heat shock proteins. FKBP52 is also thought to have a role in glutocorticoid receptor trafficking into the nucleus. FKBP36 and FKBP38 are interesting, as they do not act as PPIases. Their PPIase-like domains show the most divergence from FKBP12 when considering the residues forming the active site, Figure 1-5. FKBP36 is a glyceraldehyde-3-phosphate dehydrogenase inhibitor and is thought to have a role in spermatogenesis; it forms a complex with Hsp72 (which binds to the TPR domain) and clathrin (which binds to the PPIase-like domain). The N-terminal PPIase domain has been crystallised and shows some interesting features. Trp59 forms the base of the hydrophobic pocket in most FKBPs; in FKBP36 it is replaced with methionine, Phe46 is replaced with Arg81. FKBP38 has a much deeper narrower active site than FKBP12. Trp59 is replaced with Leu156, Phe36 with Gln138 and His87 with Arg184. Many of the aromatic residues lining the active site are replaced with leucine. Interestingly FKBP38 binds calcineurin without FK506 acting as a bridging ligand. FKBP38 has been shown to interact with Bcl-2 and Bcl-xl (mitochondrial membrane proteins important in cell viability) and the presenilins SN1 and SN2. Rheb is thought to act allosterically on FKBP38 to facilitate the binding of the C-terminal domain of FKBP38 to mTor.

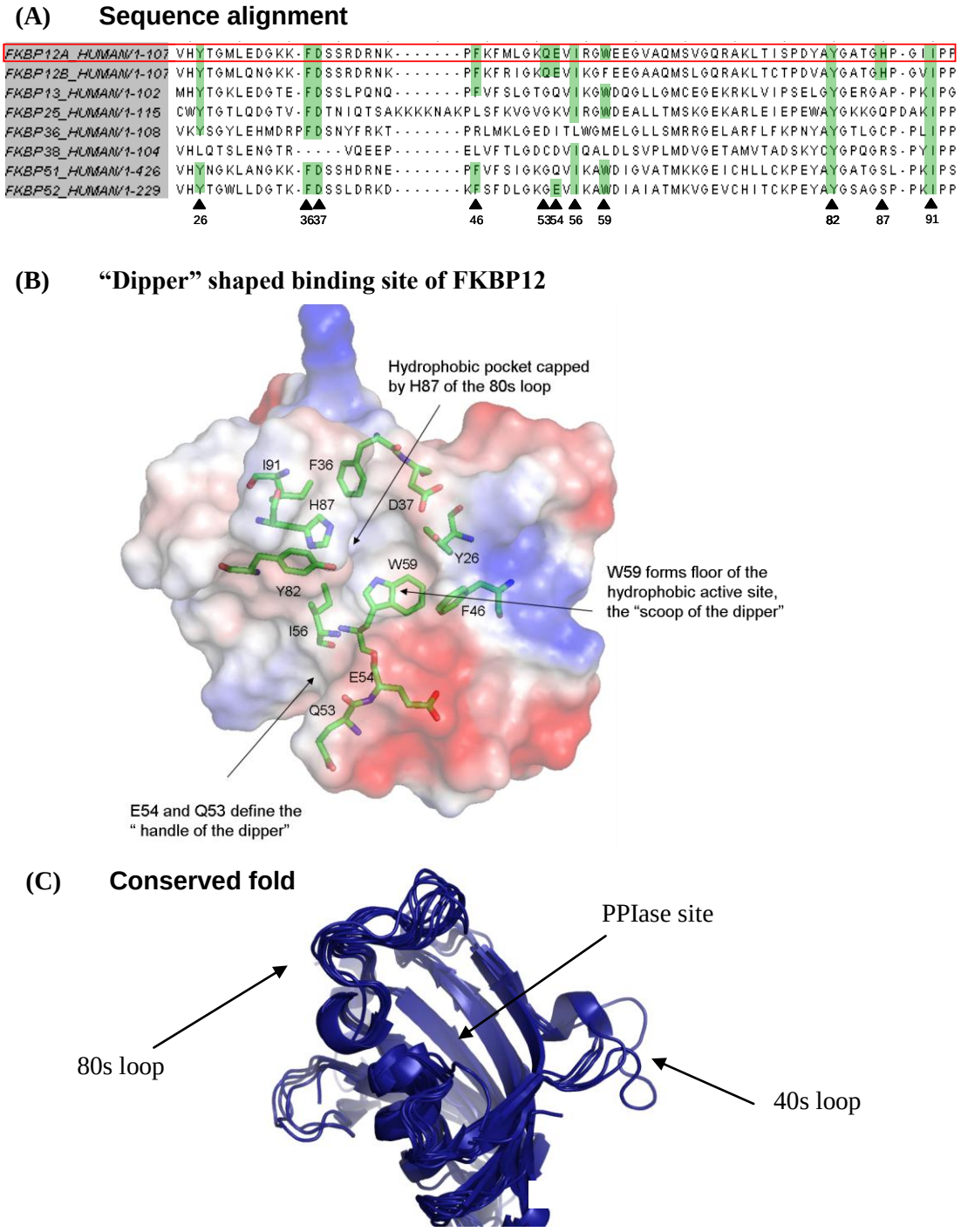


Figure 1-5 Alignment of 8 FKBP12-like PPIase domains

(A) Alignment of human FKBP PPIase domains that have structures deposited in the PDB (Table 1-2), green bars show residues in the active site that make important non-covalent interactions with rapamycin (labelled with the numbering scheme for FKBP12). (B) The active site of FKBP12. (C) FKBP12-like PPIase domains all share the same basic topology; most differences are in the flexible 80s and 40s loop regions. Structures were aligned to the PPIase domain of FKBP12 (2ppn.pdb) using PyMol (V1.02, DeLano Scientific).

The ER FKBPs form the largest, but least investigated group. The proteins all contain 2 cysteine residues that are thought to form a cystine bond and stabilise the PPIase domain; most PPIase domains contain 1 cysteine. It has been suggested that an increased stability is important in the extremes of pH present in the secretary tissue where ER FKBPs are most highly expressed (Rulten *et al.*, 2006). ER FKBPs also contain a cleavable leucine rich N-terminal sequence. FKBP22, FKBP23, FKBP60 and FKBP65 contain a calcium binding Ef-hand motif.

The nuclear FKBP25 is thought to bind to histone deacetylase through its N-terminal nuclear localisation sequence and possibly has a role in transcriptional regulation. FKBP12 has been extensively investigated and its immunosuppressive and biological roles are discussed in the next sections.

1.11.2 Isomerase activity of immunophilins

The *trans* conformation of a proline amide peptide bond (peptidyl-prolyl bond) is sterically the more favourable. However, it has been estimated that up to 30 % of prolyl peptide bonds exist in the *cis* conformation in short peptides. If the atomic structures of proteins in the PDB are examined, the percentage in the *cis* conformation is 6% and in the *trans* 94% (Wang and Etzkorn, 2006). Although, *cis* and *trans* forms of a Xaa-Pro peptide bond are almost iso-energetic, the inter-conversion through rotation has a significant energy barrier. The energy barrier to rotation is not as high as a typical amide bond but is still appreciable when compared to a C-N single bond. FKBP12 acts as a *cis-trans* isomerase, catalysing the inter-conversion by reducing the energy barrier to rotation from ~19 to ~11 kcal·mol^{-1} (Fischer, 2000; Hamelberg and McCammon, 2009).

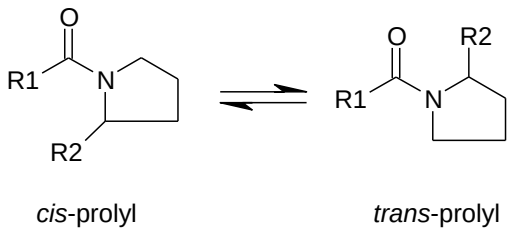


Figure 1-6 *Cis* and *trans* forms of the peptidyl-prolyl bond

The inter-conversion of the *cis* and *trans* forms of the peptidyl-prolyl peptide bond has an energy barrier of 16-22 kcal·mol^{-1} due to the partial double bond character of the peptide bond. FKBP12 catalyses the inter-conversion by reducing the energy barrier by 7.66 kcal·mol^{-1} (Fischer, 2000).

There was early interest in determining the catalytic mechanism of peptide isomerisation by FKBP12, particularly as it demonstrates a level of complexity in ligand protein interactions beyond the simplistic concept of a lock and key. It has been proposed that non-covalent interactions between FKBP12 and peptide substrate result in a transition state in which carbonyl group of the peptidyl-prolyl bond can take on a more ketone like character, the C-N bond becomes more single bond in character and hence has a lower energy barrier for rotation. A "twisted amide bond" transition state is formed and it becomes energetically favourable for the peptide to leave the binding state in a *trans* conformation (Fischer *et al.*, 1993). Rosen *et al.* examined inhibition of rotamase activity of FKBP12 by FK506; the authors concluded that FK506 acts as a mimic of the twisted amide bond seen in the transition state in the natural peptide substrate (Rosen *et al.*, 1990).

In addition changes to the isomeric state of a peptidyl-prolyl bond are also thought to act as a molecular switch in protein signalling events. FKBP12 and cyclophilin A can therefore act to regulate isomeric switching in signalling pathways (Lu *et al.*, 2007). Another interesting example of the importance of PPIase activity is cyclophilin A catalysing the *cis-trans* isomerisation of a peptidyl-prolyl bond in an HIV capsid protein that is necessary for viral repackaging (Gamble *et al.*, 1996).

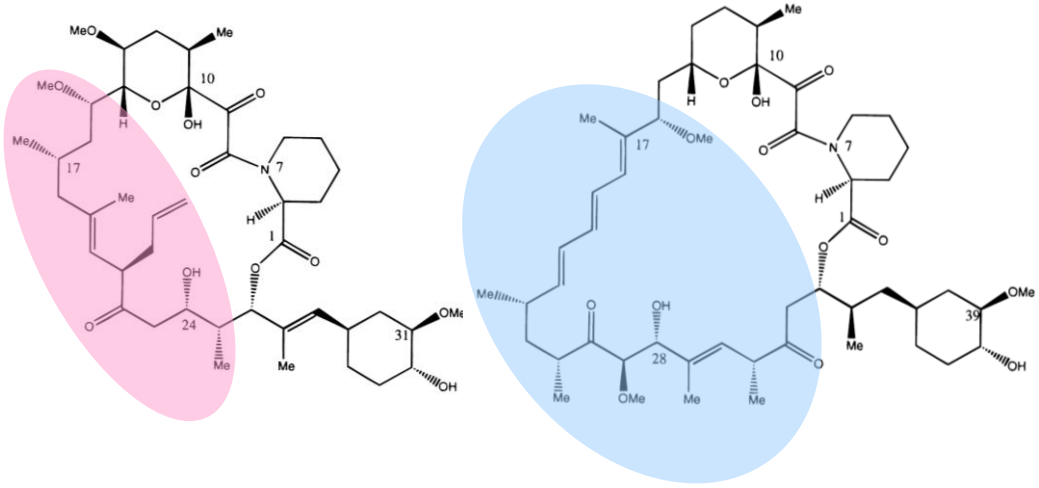
It is clear from the higher levels of mis-folded proteins seen in the production of recombinant proteins that the eukaryotic cell contains an efficient apparatus for the recognition of incorrectly folded proteins followed by repair or proteolysis (Gething and Sambrook, 1992). An incorrectly folded protein may have hydrophobic patches, which are usually buried, exposed on the surface of the protein and this can lead to aggregation. Deficiencies in the quality regulation systems of neuronal cells have been proposed as the underlying pathology in a number of neurodegenerative diseases (Gao and Hu, 2008).

1.11.3 **Biological role of FKBP12 and FKBP12.6**

The varied roles of FKBPs, beyond that of PPIase activity, may be summarised by considering the PPIase domain as a recognition domain or molecular adaptor facilitating the assembly or modulation of the activation state of a macromolecular complex. The cytoplasmic FKBPs, FKBP12 and FKBP12.6 consist of one PPIase domain and show close sequence and structural homology. FKBP12 has been shown to be involved in Ca^{2+} related signalling through interactions with ion channels (Hamilton and Steiner, 1998). FKBP12

stabilises the ryanodine receptor (RyR1) and the inositol 1,4,5-triphosphate receptor (IP_3R) improving the control of Ca^{2+} flux (Barik, 2006; Kang *et al.*, 2008). Calcineurin associates with these receptors and controls the gated channel by changes to the phosphorylation state of the complex. FKBP12 also interacts with the trans-membrane type I transforming growth factor-β (TGF-β) receptor and may mediate interactions between mast cells and neutrophils in a manner more reminiscent of a cytokine. FKBP12 stabilises the type 2 ryanodine receptor (RyR2) that is highly expressed in cardiac muscle (Deivanayagam *et al.*, 2000; Chelu *et al.*, 2004).

1.11.4 **FKBP12 and immunosuppression**



FK506 rapamycin

Figure 1-7 Structures of FK506 and rapamycin

FK506 and rapamycin share the same FKBP12 binding domain. The pink shades area indicates the portion of FK506 that binds to calcineurin. The blue shaded area indicates the portion of rapamycin that binds to mTOR. The coloured regions are known as the effector domains.

The natural antibiotic products FK506 and rapamycin were the earliest ligands of FKBP12 to be investigated. When 12000 Streptomyces strains were screened, 1 % of secondary metabolites were found to bind to FKBP12 (Driggers *et al.*, 2008). Although FK506 and rapamycin show structural similarity in the region interacting with FKBP12 their mode of action is different (Figure 1-7). They both act as chemical inducers of protein dimerisation bringing together two proteins and forming an FKBP12-drug-protein ternary complex. It is the inhibition of the second protein that modulates a cellular signalling pathway leading to the failure of T-cells to proliferate on response to antigenic stimulation. The effector loop of FK506 is solvent exposed on binding to FKBP12 and forms a compound surface with

FKBP12 that interacts with the serine/threonine phosphatase calcineurin. The phosphatase activity of calcineurin is thus inhibited. Calcineurin is important in the dephosphorylation of the transcription factors NFAT (nuclear factor of activated T-cells). Phosphorylated NFAT cannot be transported into the T-cell nucleus or bind to DNA promoting the expression of cytokines and cell surface receptors responsible for the amplification of the immune response. Rapamycin forms a ternary complex with FKBP12 and mTOR (mammalian target of rapamycin) leading to the interruption of cell cycle progression and T-cell proliferation (Abraham and Wiederrecht, 1996; Hamilton and Steiner, 1998).

Modifications to the effector domains of FKBP12 and rapamycin can change the immunosuppressive properties of the drug. For example, in the natural product ascomycin $-CH_2CH_3$ replaces $-CH_2CHCH_2$ on C21 of FK506. Ascomycin has a similar affinity for FKBP12 as FK506 but, different interactions with calcineurin due the subtle difference in the effector loop. L-685-818 (x-ray structure 1FKD.pdb represents the complex with FKBP12-L-685-818) is very similar to ascomycin with only a substitution of -H for –OH on C18 and an ethyl for an allyl group on C21 (Babine and Bender, 1997). The PPIase binding domain of the immunosuppressant drugs are discussed in the section 1.11.6 that describes FKBP12 inhibitors.

1.11.5 X-ray and NMR structures of FKBP12

There are 37 X-ray and NMR structures for FKBP12 deposited in the protein data bank (PDB, December 2008). These include structures for wild type and mutant FKBP12; free, in complex with small molecules and in FKBP12-drug-protein ternary complexes. 14 different FKBP12-small molecules complexes have been deposited in the PDB (Structures and K_i values for ligands are shown in Appendix 3).

The highest resolution x-ray structure of apo FKBP12 in the PDB is 2PNN.pdb (0.93 Å, 2008). The structure of FKBP12 is reminiscent of a 6 fingered curled hand, Figure 1-8 (A). It consists of a 5 strand anti-parallel β-sheet which can be thought of as fingers of the hand, 2 short stretches of α-helix and 3 linking loops. The 80s loop forms the thumb extending above the longer α-helix and the 40s loop a boundary to the active site. NMR NOE analysis indicates these loop regions are the most flexible in solution (Sich *et al.*, 2000). Analyses of apo and holo x-ray structures indicate that the 40s and 80s loops are stabilised upon ligand binding (Wilson *et al.*, 1995). Superposition of FKBP12 in complex with small molecules

shows most variation between protein structure in the 40's and 80's loops. Figure 1-8 (A) illustrates how the 80s loop moves in and down (folds over) on ligand binding.

The residues surrounding the base of the active site are positioned very consistently in all x-ray structures. The rim of the main pocket is lower near the top of the α-helix in a gap between the 40s and 80s loops, this leads to a channel bounded on one side by the residues Gln53 and Glu54. This channel is occupied by portions of the larger ligands, thus making the full binding cavity dipper shaped. The active site of FKBP12 is illustrated in Figure 1-5 (B) and Figure 1-8 (B). The scoop of the dipper forms the catalytically active proline binding site and has a hydrophobic base "floored" by Trp59. Phe36, Phe46 and Phe48 form the walls of the pocket. The handle of the dipper and the rim of the proline binding site contain residues that can form hydrogen bonding interactions with the larger ligands. Rapamycin makes five hydrogen bonds with the protein (Tyr26, Asp37, Gln53 Glu54, Ile56 and Tyr82). There is a hydrophobic sub-pocket in the region of His87, Ile90 and Ile91 which is relatively flexible and can accommodate various groups. Trp59 moves down (away from His87) to accommodate the larger ligands.

There have been attempts to predict the druggability of a protein based on conformational dynamics and physicochemical descriptors. A group at Abbott Laboratories has devised a druggability coefficient to assess potential drug targets (Brown and Hajduk, 2006). This coefficient aims to quantify how the effects of thermal motion and conformational changes on ligand binding may influence predictions of protein binding to small molecules. The group reports that FKBP12 has a high degree of druggability on the evidence of molecular dynamic simulations. FKBP12 appears to present an open, druggable pocket for nearly 90 % of conformations examined.

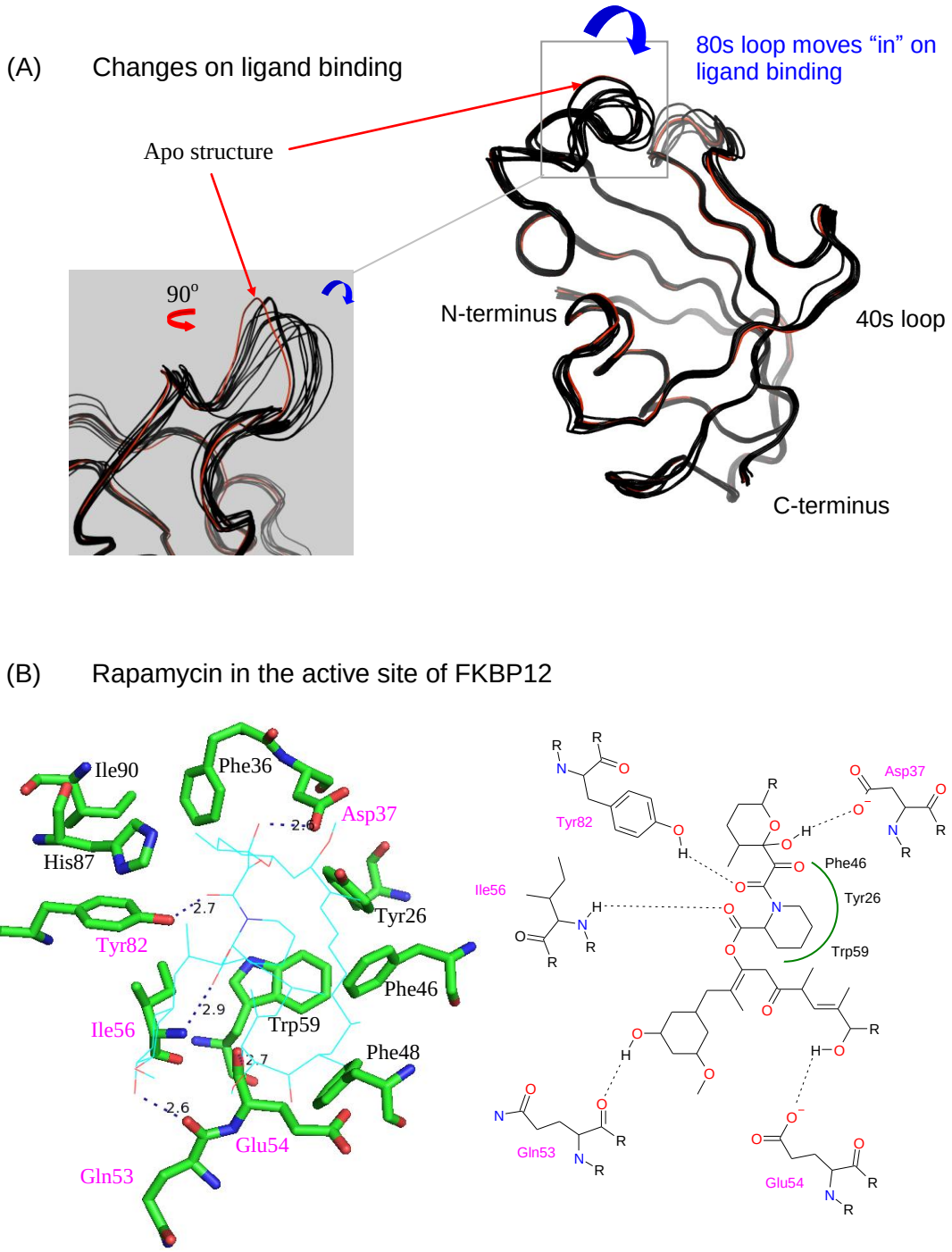


Figure 1-8 FKBP12 complexes

(A) 12 complexes of FKBP12 inhibitors were aligned with a high resolution x-ray structure of apo FKBP12 (red ribbon, 2ppn.pdb, 0.92 Å). The 80s loop is stabilised and moves in (towards the base of the active site) on ligand binding. The grey inset box shows the 80's loop rotated 90° anti-clockwise, in the vertical axis, relative to the representation of the full structure. (2PNN.pdb, 2DG3.pdb, 1FKF.pdb, 1FKH.pdb, 1FKG.pdb, 1FKI.pdb 1J4R.pdb, 1J4I.pdb 1J4H.pdb, 1F40.pdb, 1D7I.pdb, 1D7H.pdb (References and full ligand structures shown in Appendix 3). (B) Schematic diagrams of rapamycin in the active site of FKBP12. Black dashed lines to magenta labelled residues represent hydrogen bonds, labelled in Ångstrom (2DG3.pdb).

Hydrophobic and hydrogen bonding interactions for 12 of the deposited x-ray structures of protein small molecule complexes have been summarised in Table 1-3 (2 ligands have been omitted as they share a protein binding domain with rapamycin). Ligand structures (with K_i) are shown in Appendix 3. The non-covalent interactions between FKBP12 and FK506 and rapamycin have many common features due to the molecules showing structural similarity in the FKBP12 binding regions; they have K_d of 0.4 nM and 0.2 nM respectively (Bierer *et al.*, 1990). Here the non-covalent interactions of FK506 with FKBP12 are described first. The pipecolenyl ring is buried deepest in the pocket and forms hydrophobic interactions with Trp59 at the base of the pocket. This takes the place of proline in the natural substrates. Hydrogen bonding interactions are predicted with four residues. The C1 carbonyl on FK506 acts as a hydrogen bond (H-bond) acceptor to hydrogen donated by the main chain N of Ile56 (see Figure 1-7 for numbering scheme). The amide oxygen on C8 forms a H-bond acceptor for the H of Tyr82. A hydroxyl group on C10 of the pyranosyl ring forms an H-bond with Asp37 and there is a further H-bond to the main chain carbonyl of Glu54 from the C24 hydroxyl. The second carbonyl C9 of the twisted diketo moiety between the pipecolenyl ring and the pyranosyl ring forms weak hydrophobic interactions with Tyr26, Phe36 and Phe99. There are further hydrophobic interactions with Phe46 and Val55. The methyl group on C11 lies in the His87-Ile91 hydrophobic sub-pocket. In the rapamycin complex there is an additional H-bond from the C40 hydroxyl to the main chain carbonyl of Gln53 and more extensive hydrophobic interactions between the protein and cyclohexyl ring.

The 10 additional complexes are detailed in Table 1-3. They show a similar binding mode to FK506 and rapamycin. 7 of the ligands have a 5 or 6 member heterocyclic aliphatic ring that lies deepest in the pocket in a position analogous to proline in the natural substrate and the pipecolenyl ring of FK506 and rapamycin. 9/10 of the ligands make a hydrogen bonding interaction with the main chain N of Ile56 through a carbonyl group equivalent to the C1 carbonyl of FK506 that forms part of an ester linker. Interestingly the heterocyclic rings all possess a chiral center at C2. This is the S form in all x-ray structures. The same is noted for the majority of the inhibitors listed in Table 1-4. 7/10 possess a carbonyl oxygen analogous to the C8 carbonyl oxygen (FK506) that makes a hydrogen bond with the hydroxyl of Tyr82.

PDB	Ligand	Tyr26	Phe36	Asp37	Phe46	Gln53	Glu54	Val55	Ile56	Trp59	Tyr82	His87	Ile90
2DG3	rapamycin			2.61		2.63	2.67		2.93	3.45	2.66		
1FKJ	FK506			2.78			2.68		2.80	3.44	2.76		
1fkh	SBX								2.89	3.34	2.64		
1fkg	SB3								2.89	3.54	2.62		
1fki	SB1								2.87	3.47	2.75		
1j4r	FK001					2.66			2.98	3.34	2.75		
1j4i	308								2.76	3.25	2.65		
1j4h	107								2.89	3.44	2.66		
1f40	GPI1046										3.28		
1d7j	HB								2.94	3.49			
1d7i	DSS								2.85				
1d7h	DMSO								2.85				

Table 1-3 Summary of non-covalent interactions in FKBP12-inhibitor complexes

A comparison of possible hydrophilic (highlighted in blue, shortest heavy atom distance in Å) and hydrophobic (highlighted in green, shortest heavy atom distance in Å) interactions between ligands and FKBP12 from x-ray structures deposited in the PDB (distances were measured in PyMOL V1.02, DeLano Scientific). Ligand structures are illustrated in Appendix 3. Figure 1-9 shows a portion of the active site of the x-ray structures of complexes aligned with 2ppn.pdb.

It is worth noting that ligand oxygen atoms making hydrogen bonding interactions with the main chain N of Ile56 and Tyr82 occupy very similar positions to the oxygen atom of the water molecules labelled W1 and W2 seen in the apo structure of FKBP12, Figure 1-9. The carbonyl oxygen, equivalent to the C9 carbonyl of rapamycin, occupies a similar position to water molecule W3 of the apo structure. The displacement of these waters on ligand binding is predicted to provide favourable entropic contribution to free energy of binding (Connelly *et al.*, 1993). These observations suggest that Ile56, Tyr82 and Trp59 can be considered "hot spots" in the active site of FKBP12.

When docked ligand poses of similar score were visually examined during the selection of compounds for purchase the presence of the conserved non-covalent interactions between a ligand and a "hot spot" residue: Tyr82, Ile56 and Trp59 were noted. If docked poses replicated these interactions the pose represented a protein-ligand conformation where the ligand was well buried in the active site and part of the ligand mimicked proline in the natural substrate.

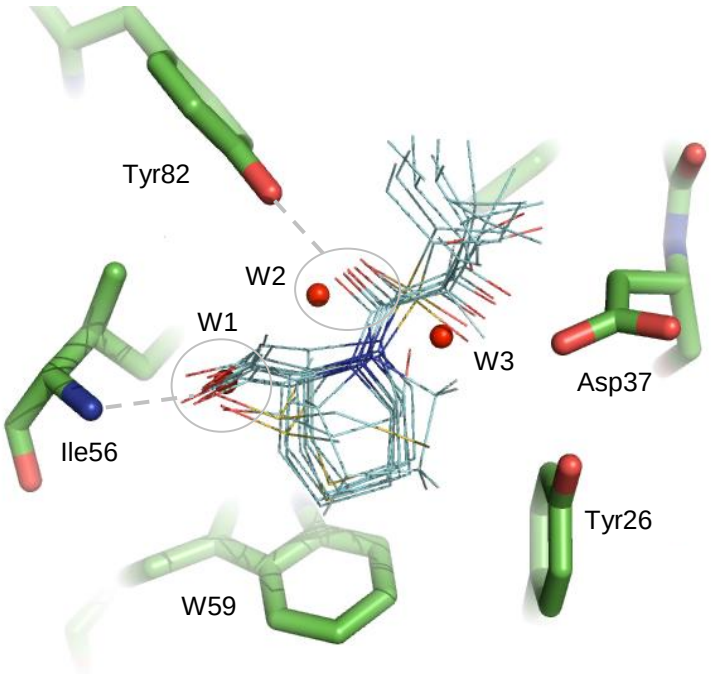


Figure 1-9 Small molecule inhibitors in complex with FKBP12

Ligand oxygen atoms making hydrogen bonding interactions with the main chain N of Ile56 and Tyr82 occupy very similar positions to the oxygen atom of the water molecules in the apo structure labelled W1 and W2. The carbonyl oxygen, equivalent to the C9 carbonyl of rapamycin, occupies a similar position to water molecule W3 (apo FKBP12, 2PNN.pdb).

1.11.6 **FKBP12 inhibitors**

FKBP12 inhibitors can be broadly divided into immunosuppressants (FK506, rapamycin and similar macrocyclic antibiotics) and inhibitors that target the PPIase domain of FKBP12, the latter group are by far the more numerous. PPIase inhibitors include compounds that have been specifically designed for the treatment of degenerative neurological conditions (Hamilton and Steiner, 1998). There is also good evidence for certain solvents binding with low affinity; this has relevance for binding assays (Burkhard *et al.*, 2000). FKBP12 inhibitors have been reviewed by Babine *et al*., , Dornan *et al*., and Wang *at al*., in some detail (Babine and Bender, 1997; Dornan *et al.*, 2003; Wang and Etzkorn, 2006). Within the constraints of a brief review this section attempts to summarise common features between inhibitors.

Many synthetic ligands are based on the FKBP12 binding domains of FK506 (binding loop) and rapamycin, Figure 1-7. Removing the effector loop and replacing it with a variety of aliphatic chains, rings or an aromatic group has proved successful. Replacement groups have been attached to the diketo moiety and ester linkage flanking the pipecolenyl ring resulting in acyclic ligands. Many of these have high affinity for FKBP12 and are potent isomerase inhibitors with K_i values for FKBP12 in the low nanomolar range, Table 1-4. Christner *et al*., (2001) produced a reduced binding model pharmacophore providing flexibility between a 5 and 6 member ring and emphasizing the importance of the α-keto amide functionality (Christner *et al.*, 2001).

For example, the pyranosyl ring in FK506 was replaced with a (dimethyl-ethyl)methyl group and the ester beyond the pipecolenyl ring linked to rings in the ligand SB3, the complex described by 1FKG.pdb (Holt *et al.*, 1993; Holt *et al.*, 1994). A 3,4,5-trimethoxy phenyl group has also been used to replace the pyranosyl ring (Armistead *et al.*, 1995). In Table 1-4 an attempt has been made to illustrate the effect of replacing key elements/functional groups in the core binding motif of FK506 and rapamycin. The core FK506/rapamycin binding motif consists of a pipecolenyl ring with a diketo linker on N7 (α-keto amide) and an ester linker on C2. A general structure is shown in Figure 1-10. Replacements to the diketo groups and ester linkage with bio-isosteres has received much attention, particularly by the groups of Hamilton and Steiner and Holt.

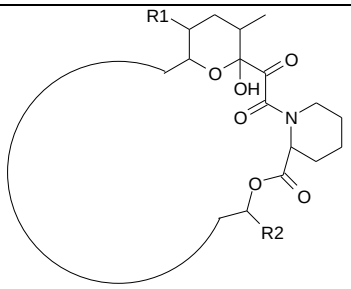
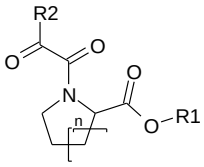
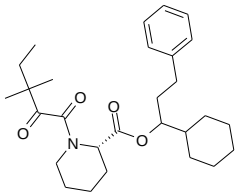
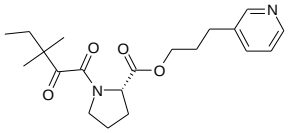
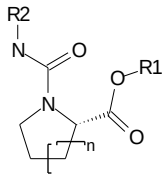
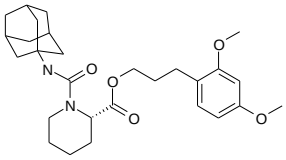
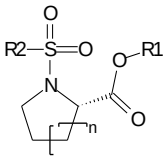
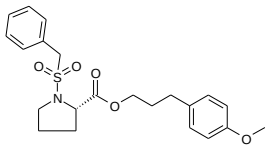
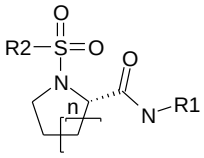
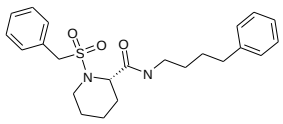
Core name	*Core Structure*	*"Best" in series, K_i*	*Reference*
M1:rapamycin core		rapamycin 0. 2 nM (Figure 1-7)	(Bierer *et al.*, 1990)
M2 (non-cyclic)		 SB3 7 nM	(Holt *et al.*, 1993; Holt *et al.*, 1994)
		 GPI-1046 7.5 nM	(Sich *et al.*, 2000)
M3 (α-keto amide replaced with urea/thiourea)		 12 nM	(Choi *et al.*, 2002)
M4 (α-keto amide replaced with sulphonamide)		 22 nM	(Choi *et al.*, 2002)
M5 (α-keto amide replaced with sulphonamide; ester with amide)		 110 nM	(Wei *et al.*, 2002)

Table 1-4 Core structures FKBP12 inhibitors

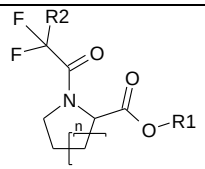
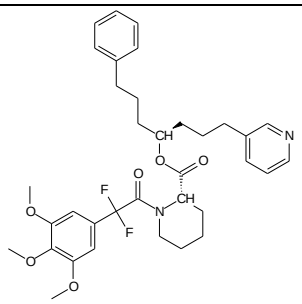
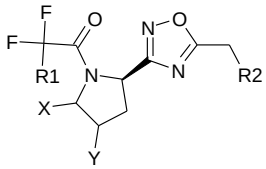
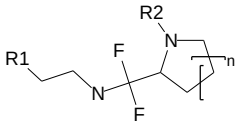
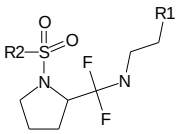
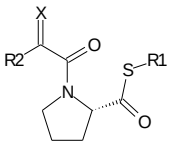
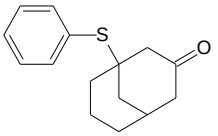
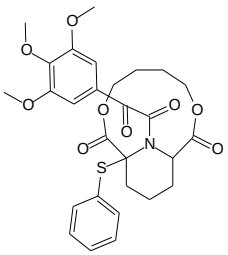
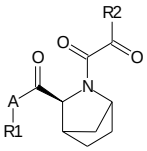
Core name	*Core Structure*	*"Best" in series, K_i*	*Reference*
M6 (C9 carbonyl replaced with 2-aryl-2,2-difluoroacetamide)		 FK001 19 nM	(Dubowchik *et al.*, 2001)
M7 (C9 carbonyl replaced with 2-aryl-2,2-difluoroacetamide)			Patent: Taguchi, Minoru; Wataya, Kengo. (2004)
M8 (α-keto amide replaced with (difluoromethyl)amine)			Patent: Watatani, Kengo; Taguchi, Minoru. (2004)
M9 (α-keto amide replaced sulphonamide; ester with (difluoromethyl)amine			Patent: Watatani, Kengo; Taguchi, Minoru. (2004)
M10 (ester replaced with thioester)			Patent 6376517
M11		9200 nM	(Limburg *et al.*, 2003)
M12		280 nM	(Limburg *et al.*, 2003)
M13		IC50≥670 nM	(Limburg *et al.*, 2003)

Table 1-4 Core structures FKBP12 inhibitors (Part 2)

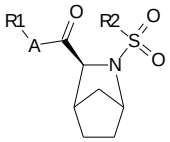
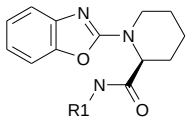
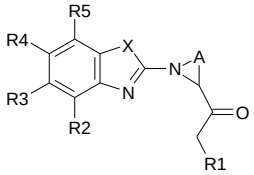
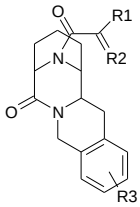
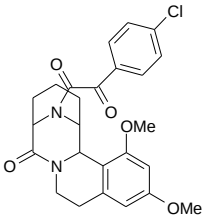
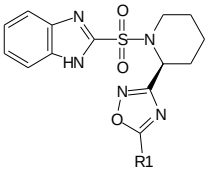
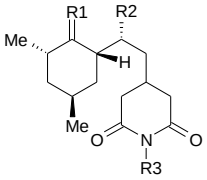
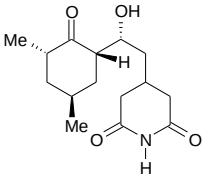
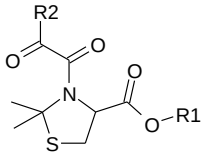
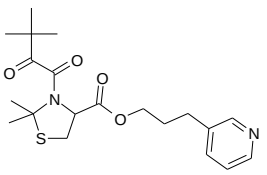
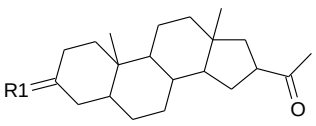
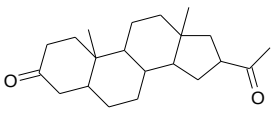
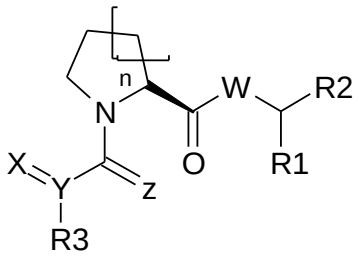
Core name	*Core Structure*	*"Best" in series, K_i*	*Reference*
M14		IC50≥650 nM	(Limburg *et al.*, 2003)
M15			Kemp *et al.* (Pfizer). Patent number: 6372736
M16			Kemp *et al.* (Pfizer). Patent number: 6372736
M17		 72 nM	(Hudack, Jr. *et al.*, 2006)
M18			Bull. Patent (Pfizer)
M19		 3400 nM	(Christner *et al.*, 1999; Edlich *et al.*, 2006)
M20			(Zhao *et al.*, 2006)
M21*****		 ≥ 7000 nM	(Burkhard *et al.*, 1999)

Table 1-4 Core structures FKBP12 inhibitors (Part 3)

***** This compound was discovered using the docking program SANDOCK.

A simplistic summary of these modification strategies might be: substitutions to the linkers to the pipecolenyl ring aims to reinforce H-bond interactions; principally with Ile56, Tyr82 and Glu54. Addition of aromatic hydrophobic groups in place of the pyranose or cyclohexyl rings aims to contribute to hydrophobic interactions, particularly in the hydrophobic pocket adjacent to His87 (pyranose ring replacement). Table 1-4 shows how it is possible to substitute in the linker regions without too much of a decrease in affinity for FKBP12 (cores M1-10). Choi *et al.* found that including a hydrophobic group on a 3-4 carbon linker to the C1 ester improved activity (Choi *et al.*, 2002). The length of the linker presumably facililitates positioning of a group analogous to the cyclohexyl ring in the hydrophobic trench adjacent to Gln53 and Glu54. Appendix 3 includes two examples of a series of ligands in which the C8 carbon of the diketo moiety of FK506 has been replaced by sulfur (M5; 1J4H.pdb and 1J4I.pdb, (Sun *et al.*, 2003)). Many core structures have been filed as Patents; for these examples affinity data is not always available.



X =H, O, CH_2, CH_3, F_2

Y =C, N, S

Z =H, N, O

W=N, O

R1=H, Me, Ph, $(CH_2)_n$Ph, $COOCMe_3$

R2=H, $CHMe_2$, CH_2CH_2Ph, $(CH_2)_3$-3-Pyr

R3=c-hexyl, CMe_2Et, 3,4,5MeO_3Ph,

n = 1, 2

Note: some element changes result in a change in C hybridisation from Sp_2 to Sp_3, appropriate addition of H is assumed

Figure 1-10 Core binding motif of many FKBP12 inhibitors

Summary structure of acyclic derivatives of FK506/rapamycin and some examples of substituted elements/groups investigated in the literature.

An inhibitor with a ring containing sulfur and nitrogen (M20) has been reported (Zhao *et al.*, 2006). The top hit in this series is very similar to compound GPI-1046 (M2) apart from the heterocyclic ring mimicking proline in the active site.

Inhibitors with a scaffold not based around the pipecolenyl ring are less numerous and, as a generalisation, of lower affinity. There has been some interest in more rigid inhibitors, such as cyclic urea derivatives, the aim being to produce better mimic the twisted intermediate implicated in the model of PPIase action (Dragovich *et al.*, 1996). Burkhard *et al*., (1999) discovered some novel inhibitors using the molecular docking program SANDOCK (Sandoz Pharmaceticals, Basel, Switzerland). These include the steroids 5β-pregnan-3-20-dione, 5β-pregnan-3α-ol-20-one and benzyloxycarbonyl–L-proline derivatives (Burkhard *et al.*, 1999). Other interesting compounds are the bicyclic ring system reported by Limburg *et al*. (2003)

and the polycyclic aza-amide compounds described by Hudack *et al*., (2006) (Limburg *et al.*, 2003; Hudack, Jr. *et al.*, 2006).

There has been relatively little published on novel FKBP12 inhibitors in the last two years. There have been two studies reporting fragment based approaches. San Diego Centre for Chemical Genomics carried out a high-throughput NMR screen of 3775 fragment-like compounds (Stebbins *et al.*, 2007). Section 4.4.2 describes structure activity relationships for the most active cluster of compounds. The active compound 2-[(4-methylphenyl)sulfonyl]-1-(morpholin-4-yl)ethan-1-one (SD1) has been used as a medium affinity positive control in this study. SD1 has a K_d of 6.8 μM for FKBP12 but is around 100x more soluble in aqueous buffer than the low nanomolar affinity immunosuppressive drugs. A study by Rohrig *et al.* builds on the SAR by NMR studies of Fesiks' group and the work on linkers by Armistead (Armistead *et al.*, 1995; Shuker *et al.*, 1996; Rohrig *et al.*, 2007). Rohrigs' study does not offer any novel scaffolds to replace the core motif of rapamycin.

1.11.7 **FKBP12 inhibitors and neurological conditions**

It was found that FKBP12 has an up to 40 % higher level of expression in the rat brain compared to cells of the immune system. This led to supposition that FKBP12 is involved in neurological function. Levels were particularly high in regions of the brain that have implications for Parkinson's and Alzheimer's disease. It has been reported that FK506 can enhance neurite out growth in cultures of rat PC12 pheochromaocytoma cells and sensory ganglia and FKBP12 inhibitors may be useful in cases of neuronal damage (Lyons *et al.*, 1994). Other studies showed a neuroprotective role for FK506 in a model for cerebral ischemia, a condition observed in stroke patients. This was explained by interaction of FK506 with the calcium-calmodulin dependent enzyme nitric oxide synthase (NOS) by blocking the action of calcineurin and keeping NOS in a hyperphosphorylated state (Hamilton and Steiner, 1998; Brecht *et al.*, 2003). Moreover, neurotrophic effects were observed for non-immunosuppressive inhibitors of the PPIase domain of FKBP12 that do not bind to calcineurin or mTor (Edlich *et al.*, 2006). The neurotrophic effect is therefore not exclusively dependent on ligand binding to calcineurin. The mechanism of neurotrophic effects has not been fully explained. The understanding is complicated by the presence of PPIase domains in all members of the FKBP family of proteins. It has also been argued that the rotamase function of FKBP12 rescues mis-folded proteins and can in certain circumstances lead to the formation of undesirable protein plaques in the brain.

1.12 Project outlines

1.12.1 Chapter 2: Protein Production

The six histidine N-terminal construct of FKBP12 was over expressed in *E.coli* and purified to greater than 95 % purity. The protein was biochemically and biophysically characterised prior to experimental ligand screening experiments.

1.12.2 Chapter 3: ESI-MS in drug discovery

An aim of this part of the study was to find the correct solution and instrument parameters for flying the six histidine N-terminal construct of FKBP12 and preserve protein in complex with ligand. The project examined correlations between rank orders of ligand affinity determined in ESI-MS and experiments carried out in solution using a test set of structurally diverse inhibitors in the nano to millimolar range.

1.12.3 Chapter 4: Hydrophobic effects versus electrostatic interactions in ESI-MS?

A series of experiments tested the hypothesis: *complexes with a significant enthalpic contribution to the free energy of binding arising from electrostatic interactions dissociate less in electrospray*. One might make the assumption that the active site of a protein contains amino acids complementary to physiologically important binding partners; in respect to both charge and hydrophobicity. Does it therefore follow that some proteins may be more amenable to screening using ESI-MS than others? This study compared the behaviour of FKBP12 in electrospray with two protein-ligand complexes where electrostatic contributions to binding provide the dominant energy term. VC50 (cone voltage required to dissociate 50 % of the complex) were compared to K_d and thermodynamic profiles determined by isothermal titration calorimetry.

1.12.4 Chapter 5: The Spectral Ghost

During the initial experimentation it became apparent that preserving complex for medium affinity ligands (1 μM to 100 μM) of FKBP12 was challenging. This raised practical difficulties, the majority of the compounds in the test set had affinity in the micromolar range. High/medium affinity ligands are also much less likely than low affinity ligands to be picked up in the early stages of the drug discovery process. However, careful examination of the electrospray spectrum, with and without the presence of a ligand, showed changes to the charge-state distribution of the protein and the possibility of using charge-state shift as a sensitive tool to screen for weak non-covalent interactions.

1.12.5 Chapter 6: Biophysical and biochemical ligand binding studies of FKBP12

The ligand test set was used to compare results from ESI-MS with other biophysical and biochemical techniques.

1.12.6 Chapter 7: Screening for Inhibitors of FKBP12

X-ray structures of FKBP12 were used as docking templates for virtual screening experiments using the programs LIDAEUS (Wu *et al.*, 2003) and FlexX (Rarey *et al.*, 1996). 3D structures of ligands were used by UFSRAT, a ligand similarity based method, where known inhibitors of FKBP12 were compared to molecules in a virtual library (Steven Shave, Edinburgh University). A Bayesian machine learning method was used to assign a relative activity probability to a compound in a library (Angelopoulos *et al.*, 2009). Compounds from four libraries were screened against FKBP12; EDULISS 2 (3.8 M unique compounds from commercial and academic sources (Hsin, 2009) ; a subset of the Specs catalogue (15 000, commercially available compounds. Specs, Delft, The Netherlands); a fragment library (96 compounds selected from The Maybridge Ro3 collection, UK) and the KB series (a library designed to target immunophilins).

1.12.7 Chapter 8: Human Cyclophilin A – Testing the KB series

The KB series of compounds were tested against the immunophilin cyclophilin A by biophysical assay and x-ray crystallography. This enzyme is structurally different to FKBP12 but does exhibit peptidyl-prolyl isomerase activity. Crystals of hcypA were grown and then soaked in saturated solutions of potential inhibitors. This method has proved useful in characterising the binding of very weak inhibitors.

2 Chapter 2 Protein Production

2.1 Introduction

Recombinant human N-terminal hexa-histine tagged FKBP12 (FKBP12, Figure 2-1) was expressed in *E.coli* and purified under native conditions following a modified version of the protocol described by Wear *et al.*, 2007b. Pure protein yields were initially in the order of 6 mg·L^{-1} culture. These were improved 1.5 fold by the use of a 10 L fermentor (Bioflo 4500 fermentor, New Brunswick Scientific) and cell disruption (Constant Cell Disruption Systems). Cell disruption was found to give more consistent cell lysis than sonification. Modification to the affinity chromatography protocol increased the purity of the sample after step one of the purification and reduced losses in the concentration steps. Final protein yields were in the region of 14 mg·L^{-1} culture.

```
        10         20         30         40         50         60
MRGSHHHHHH GSMGVQVETI SPGDGRTFPK RGQTCVVHYT GMLEDGKKFD SSRDRNKPFK

        70         80         90        100        110        120
FMLGKQEVIR GWEEGVAQMS VGQRAKLTIS PDYAYGATGH PGIIPPHATL VFDVELLKLE
```

Figure 2-1 Sequence of recombinant human N-terminal hexa-histidine tagged FKBP12.

Sequence in black is that of wild type human FKBP12, sequence in blue was added to allow an efficient, high yield purification by Ni^{2+}-affinity chromatography.

FKBP12 was judged to be greater than 95 % purity by SDS-PAGE (Laemmli, 1970). The protein was active as a peptidyl-prolyl isomerase with a turnover number (K_{cat}) of 300 ± 40 s^{-1} for the preferred peptide substrate N-succinyl-Ala-Leu-Pro-Phe-p-nitroaniline; this is in agreement with accepted values (Kofron *et al.*, 1991). Recombinant FKBP12 exhibited an affinity for rapamycin in agreement with the literature (Bierer *et al.*, 1990; Connelly and Thomson, 1992; Wear and Walkinshaw, 2007; Wear *et al.*, 2007b).

2.2 FKBP12 Expression and Purification

2.2.1 Transformation and overexpression of FKBP12 in *E.coli*

The plasmid for recombinant FKBP12 was constructed by Dr. Alan Paterson and consisted of an open reading frame for FKBP12 inserted and ligated into a pQE-30 vector via the HindIII and BamHI sites (Wear *et al.*, 2007b). Competent *E.coli* cells (BL21 Star^{TM}, Novagen) were transformed using the manufacturer's standard protocol. Transformed cells were grown from a single colony in two times TY media, 100 μg·ml^{-1} carbenicillin at 37 ^{o}C with 260 rpm shaking, to an Absorbance$_{600nm}$ ~ 0.6. Over expression of FKBP12 was induced by addition of 1 mM IPTG and cells harvested after a further 4 hours (Figure 2-2). Protein yields were increased around 1.5 fold in a 10 L Bioflo 4500 fermentor (New Brunswick Scientific) compared to growth in 2 L conical flasks in 500 ml batches. It is believed that increased levels of cell mass obtained in fermentor runs are in part attributable to higher levels of aeration (Bauer and Shiloach, 1974).

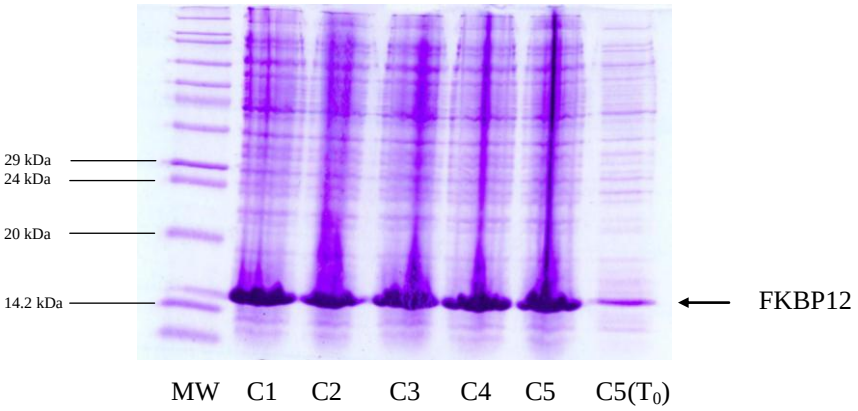


Figure 2-2 Overexpression of 6H-FKBP12.

15 % acrylamide SDS-PAGE denaturing gel showing overexpression of FKBP12 in *E.coli* (recombinant strain BL21star^{TM}-pQE30-6H-FKBP12) grown in 2xTY media at 37 ^{o}C with 260 rpm shaking. Cells were induced after reaching an optical density of A_{600nm} ~ 0.6 with 1 mM IPTG. Lanes C1 to C4 show soluble fraction from four separate colonies, grown in 500 ml batches, after 4 hours induction with IPTG. Lane C5 shows whole cell extract from colony 5 grown under similar conditions in a 10 L fermentor after 4 hours induction with IPTG and lane C5(T0) the sample C5 prior to induction. FKBP12 runs at 14.6 kDa in SDS-PAGE (15 % acrylamide).The gel was stained with Coomassie blue stain (Brilliant blue R, Sigma-Aldrich). Molecular weight marker (MW) Sigma wide range (Sigma-Aldrich).

Cell cultures were centrifuged at 50000 g for 30 minutes at 4 ^{o}C; the pellet was frozen in liquid nitrogen and stored at -80 ^{o}C until required. Active protein was successfully purified from cell pellets up to 1 year old.

2.2.2 **Affinity chromatography**

Cell pellets were rapidly defrosted and re-suspended in ice cold PBS, pH 7.5; 1 mM sodium azide; 0.1 % TX-100; 100 mM PMSF; 100 μM benzamidine; 5 mM β-mercapto ethanol; protease inhibitors (2 mini protease inhibitor cocktail tablets per 1 L culture, Roche); (30 $ml \cdot L^{-1}$ of original culture). All further purification steps were carried out at 4 ^{o}C. Cells were lysed by single pass cell disruption (Constant Cell Disruption Systems) at 25 psi. The cell extract was centrifuged at 50000 g for 1 hour, the supernatant collected, filtered to 0.2 μm and adjusted to 20 mM imidazole by the addition of a concentrated imidazole solution buffered to pH 7.5. The extract was passed over a HisPrep FF 16/10 nickel sepharose column (20 ml, GE Healthcare) using Acta-FPLC (GE Healthcare) equipment at a flow rate of 0.7 $ml \cdot min^{-1}$. The column was washed using at least 4 column volumes of PBS, pH 7.5; 1 mM sodium azide; 5 mM β-mercaptoethanol. The FKBP12 was eluted with a gradient of 4 - 100 % 500 mM imidazole, pH 7.5 over at least 15 column volumes at a flow rate of 2.5 $ml \cdot min^{-1}$. The protein elution peaked at 330 mM imidazole (Figure 2-3).

It became necessary to modify the initial two-step purification protocol after the increased protein yields afforded by the use of the fermentor and cell disruption apparatus. High levels of FKBP12 expression were accompanied by expression of an impurity of relatively low solubility at 4 ^{o}C. This impurity bound non-specifically to the nickel affinity column. This impurity elution peak had a significant overlap with the elution peak for FKBP12. Precipitation of impurities during sample concentration prior to gel-filtration pulled down a proportion of FKBP12, thus reducing yields. This problem was addressed in two ways. Raw cell extract was loaded onto the Ni^{2+}-affinity column with the addition of a low concentration of imidazole (20 mM). This reduced non-specific binding. In addition, use of a 20 ml Ni^{2+} affinity column gave some chromatographic separation between the eluted impurity and FKBP12. 1 ml fractions were collected and analysed by SDS-PAGE (Laemmli, 1970). Fractions containing FKBP12 were pooled and concentrated by spin concentrator (Vivaspin, exclusion limit 5 kDa) prior to gel filtration.

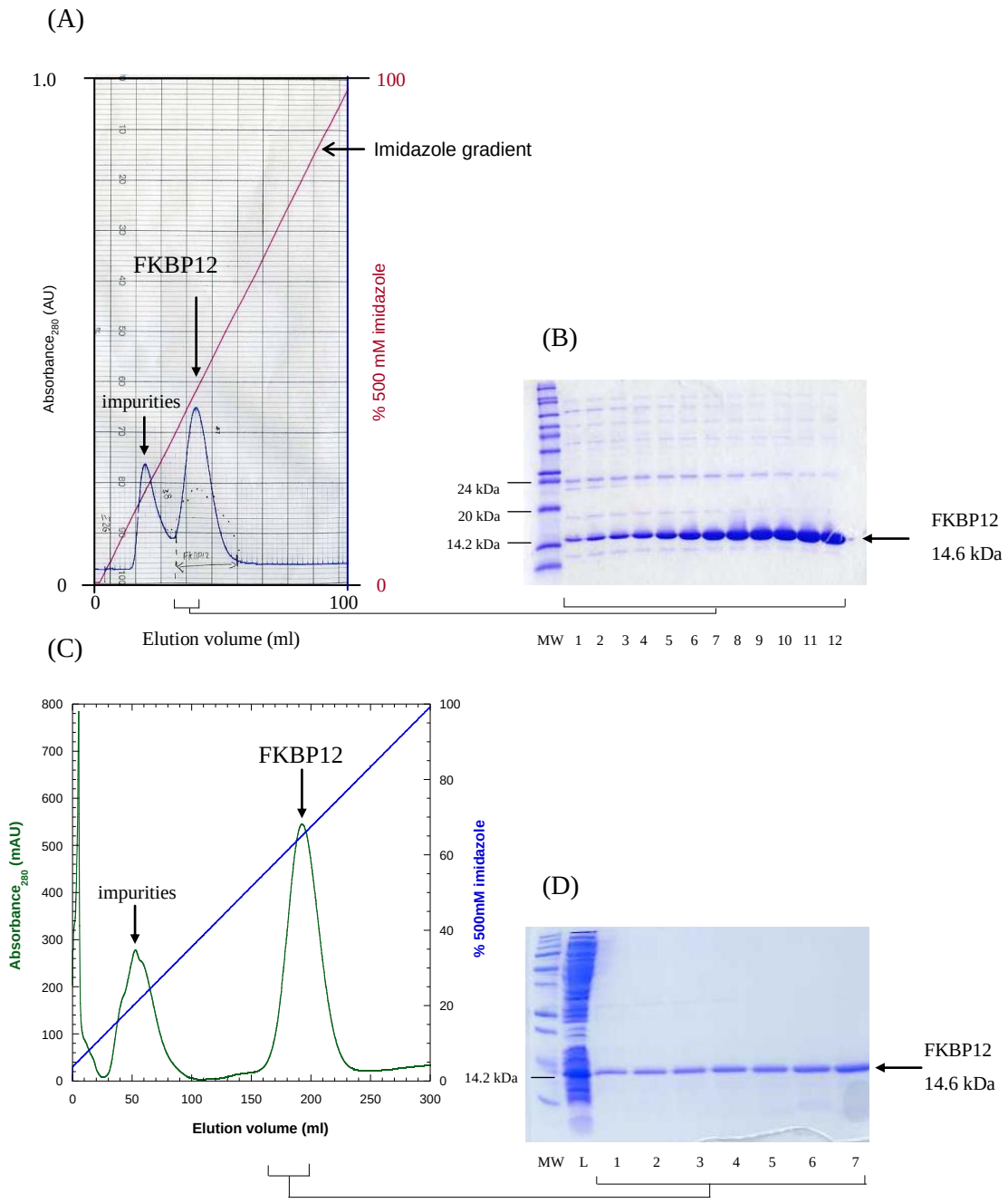


Figure 2-3 Nickel affinity Chromatography.

Use of a larger volume column and modified protocol removed a greater proportion of the impurities at the first step of purification. (A) Elution profile for FKBP12 using a 5 ml Ni^{2+}-NTA column (Superflow resin, Quiagen) eluted by imidazole gradient (0-100% 500 mM imidazole over 20 column volumes). (B) 15 % acrylamide SDS-PAGE gel showing FKBP12 after the 1st step of purification using a 5 ml Ni^{2+}-NTA column. (C) Elution profile for FKBP12 using a 20 ml HisPrep FF 16/10 column (GE Healthcare) eluted with 4 - 100 % 500mM imidazole over 15 column volumes. (D) 4 - 15 % gradient SDS-PAGE gel showing FKBP12 after the 1st step of purification using a 20 ml HisPrep column. Molecular weight marker (MW) Sigma wide range, lane L is the sample loaded on the column. Gels were stained with Coomassie blue stain.

2.2.3 **Size exclusion chromatography**

An S200 HiPrep gel filtration column (120 ml, GE Healthcare) was pre-equilibrated with PBS, pH 7.5; 1 mM DTT; 0.5 mM EDTA, and loaded with a concentrated sample of the protein from the 1st stage of purification. The protein was loaded and eluted at a flow rate of 0.5 ml·min^{-1}. The majority of the sample eluted corresponded to monomeric FKBP12. FKBP12 fractions were pooled and concentrated prior to exchange into the appropriate experimental buffer. Buffer exchange was carried out manually on PD10 rapid desalt columns (GE Healthcare) or on an Acta FPLC system using HiPrep Desalt columns (26/10, GE Healthcare) and a flow rate of 5 ml·min^{-1}.

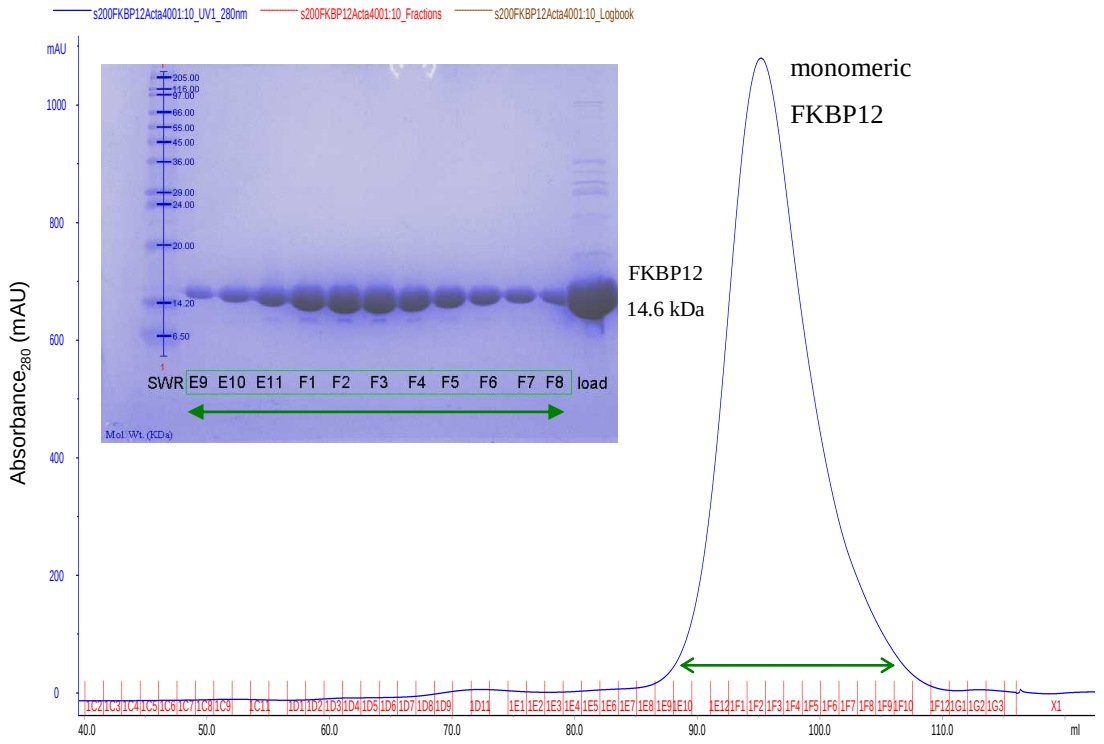


Figure 2-4 Size exclusion chromatography.

Elution profile from a S200 HiPrep gel filtration column (GE Healthcare) showing a dominant peak corresponding to monomeric FKBP12. Eluted with PBS, pH 7.4; 1 mM DTT; 0.5 mM EDTA; at a flow rate of 0.5 ml·min^{-1}. Fractions E9 to F8 were pooled; shown by the green double-headed arrow. Molecular weight marker (MW) Sigma wide range, labelled in kDa. Gels were stained with Coomassie brilliant blue stain.

2.3 Physical and Biochemical characterisation of FKBP12

2.3.1 SDS-PAGE

Protein samples were assessed for purity by 15 % acrylamide SDS-PAGE on the Mini-PROTEAN Tetra Cell system (Bio-Rad). Proteins were stained using Coomassie Brilliant Blue R (Sigma-Aldrich). Protein purity and molecular weight was estimated from band intensity and location using a Molecular Imager Gel Doc XR gel scanner and Quantity One V4.6.5 software (Bio-Rad). Figure 2-5 shows FKBP12 purity to be greater than 95 % for the pooled fractions as determined by band densitometry.

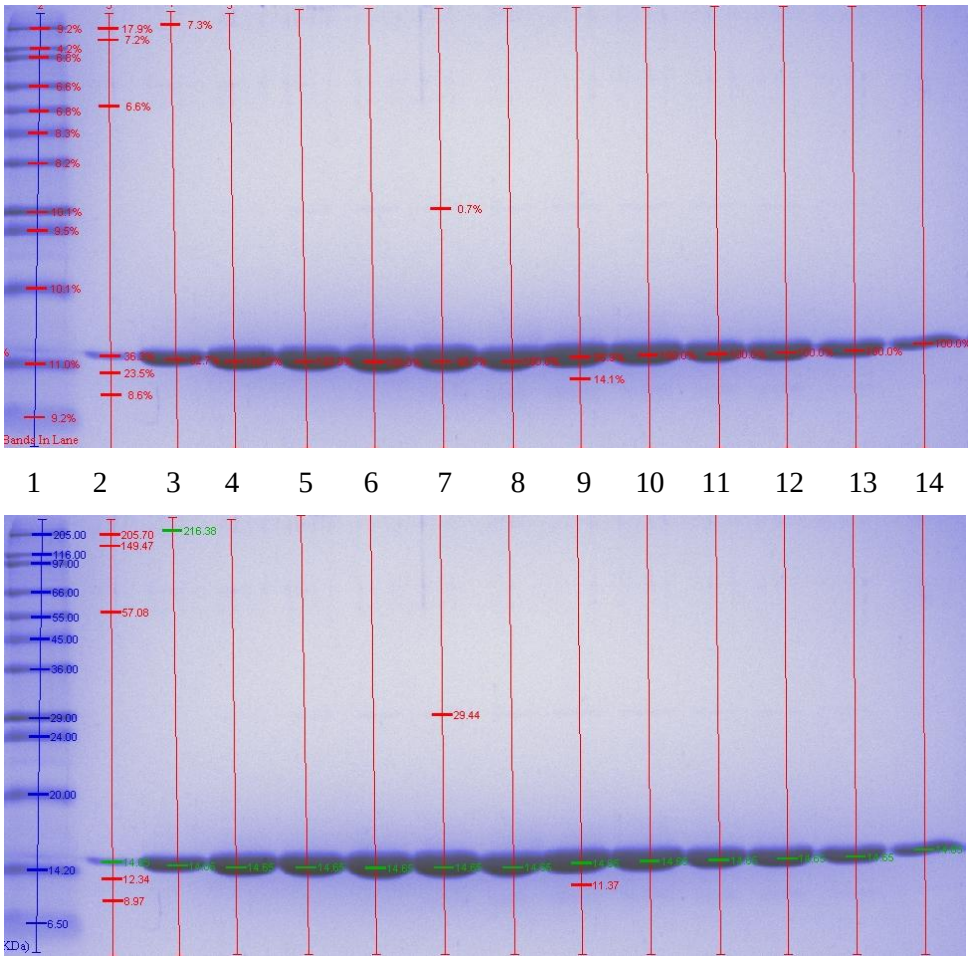


Figure 2-5 Assessment of protein purity by SDS-Page band densitometry.

Representative 15 % SDS-PAGE denaturing gel, FKBP12 in 50 mM ammonium acetate, pH 6.8. Lane 1 molecular, weight marker Sigma wide range. Lanes 2-14 elution fractions from rapid desalt column (HiPrep 26/10, GE Healthcare). (A) Densitometric analysis; red bars show % intensity of each band by lane. Fraction in lane 1 was excluded from pooling. (B) Marker bands are labelled with blue bars (kDa). Lanes 2-14 bands are labelled with red bars for impurities and green bars for FKBP12. FKBP12 ran at 14.65 kDa relative to the molecular weight standards. The gels were stained with Coomassie blue stain.

Protein was stored on ice for up to 6 weeks in a buffered solution, pH 7.4. SDS-PAGE showed a small proportion (~ 5 %) of the sample to have a mass less than that of FKBP12 after ~ 8 weeks. FKBP12 showed slightly less breakdown in 50 mM ammonium acetate at pH 6.8 than PBS at PH 7.4.

2.3.2 Determination of Protein Concentration

Protein concentration of FKBP12 was routinely determined from absorption at 280 nm (Jasco V-550 spectrometer) using an extinction coefficient of 9970 $M^{-1} \cdot cm^{-1}$. The extinction coefficient was calculated using the ExPASy ProtParam server (www.expasy.org/tools/protparam.html, June 2008), (Gasteiger *et al.*, 2005)). Protein concentration was also determined by the bicinchoninic acid (BCA) assay with bovine serum albumin standard (Brown *et al.*, 1989). The BCA colour change relies on the reduction of copper. Therefore, any reducing agents will contribute to the colour change. FKBP12 and the protein standard were precipitated and re-suspended to remove possible interfering substances prior to the assay.

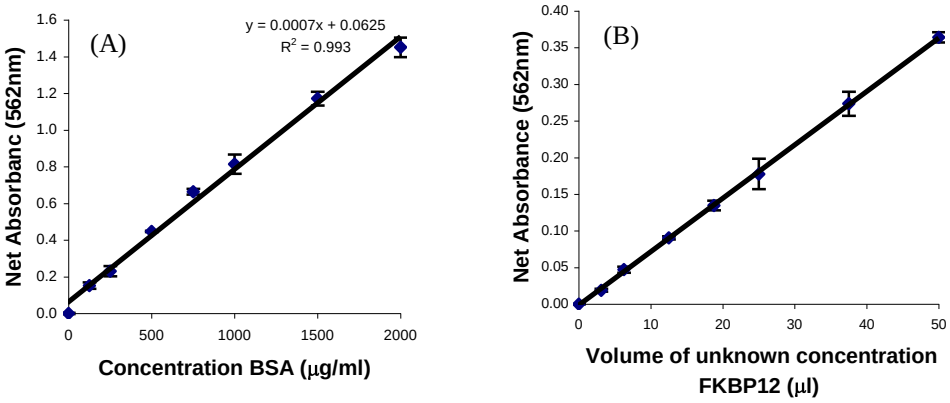


Figure 2-6 BCA assay calibration curves

(A) Net absorbance as a function of BSA concentration. (B) Absorbance of a serial dilution of an unknown concentration of FKBP12. All results are an average of 3 experiments, standard errors are shown.

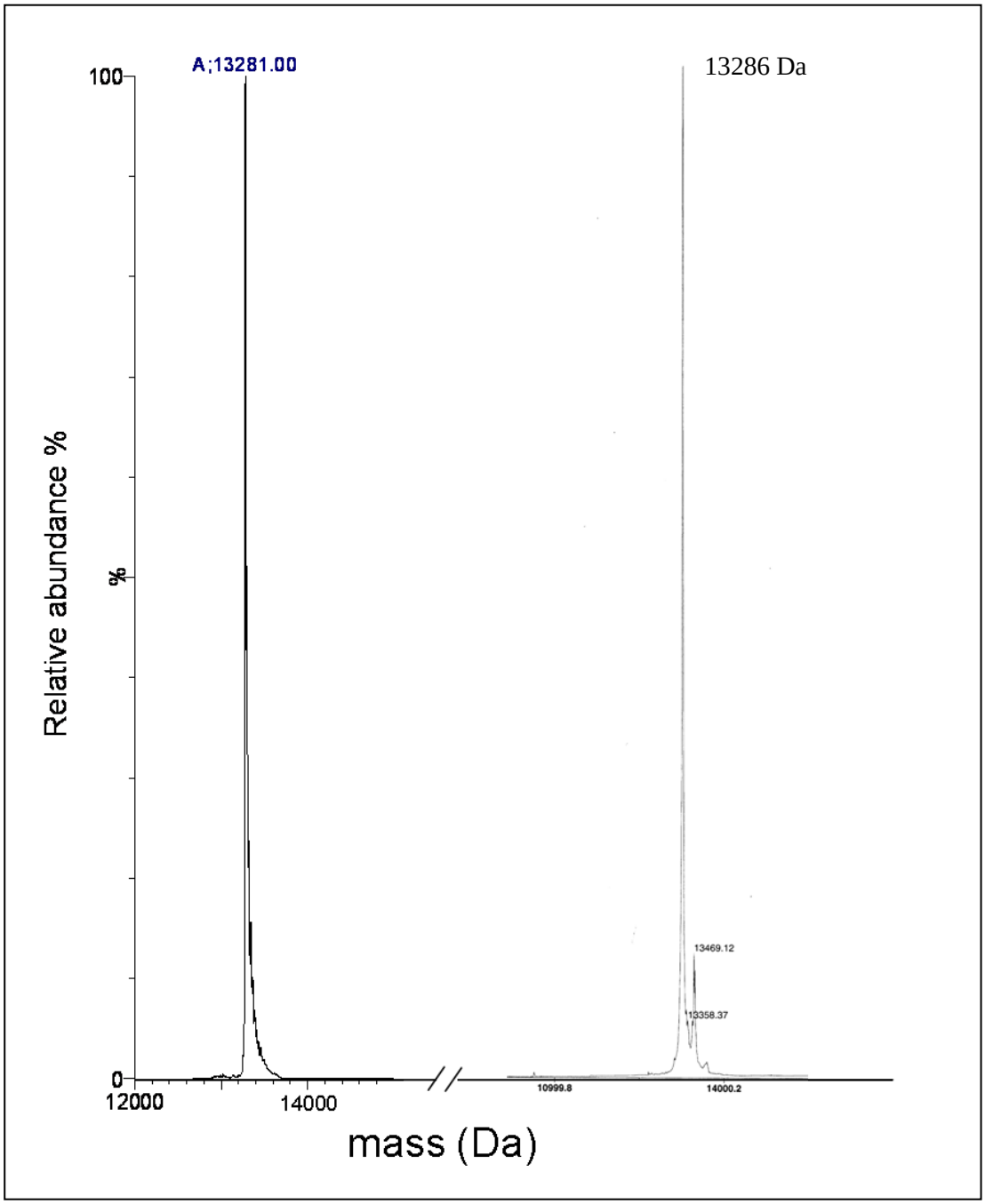
Figure 2-6 shows a typical BSA (Bovine serum albumin) calibration curve and absorption of a serial dilution of an unknown concentration of FKBP12. Measurement by absorption at 280 nM was within 5 % of that calculated by BCA.

2.3.3 **ESI-MS and MALDI Mass Spectrometry**

The MALDI mass spectrum of FKBP12 was recorded on a Voyager DE-STR MALDI-TOF (Applied Biosystems) instrument using a sinapinic acid matrix (0.5 µl of 2 $mg \cdot ml^{-1}$ FKBP12, 0.5 µl of matrix) The ESI-MS analysis was performed as described in 3.6.1.

The MALDI spectrum was dominated by a monomeric peak corresponding to 13286 a, Figure 2-7. There was an additional signal at 13469 Da, this formed less than 5 % of the total observed intensity. The mass of the dominant peak is consistent with that measured in ESI-MS (13282±4 Da. The calculated average mass of FKBP12 is 13349.1 Da and the monoisotopic mass 13340.7 Da (www.expasy.org/tools/protparam.html, June 2008), (Gasteiger *et al.*, 2005)) The majority of the sample (> 95 %) is within 1 % of the calculated value. The difference between the calculated value and the experimental value is around 80 Da. It has been suggested that post translational N-terminal deletions are a source of mass discrepancies in MS for recombinant proteins expressed in *E.coli*. However, the observed data is inconsistent with methionine deletion. Additionally, methionine aminopeptidase from *E.coli* does not usually show substrate specificity for sequences with arginine following methionine at the N-terminus (Frottin *et al.*, 2006).

The 13469.12 Da mass component (Δm = 183 Da) seen on the MALDI spectrum (B) was seen in ESI-MS for some protein batches, there was considerable variability between grow-ups; it was always less than 5 % of the signal. The presence of the species in ESI-MS and MALDI spectra suggests that the increase in mass is not due to a sinapinic acid adduct (MW = 224 Da). The illustrated ESI-MS spectrum, (A), exhibits no clear peak above the envelope at this mass. We considered the possibility that this signal was due to plasticizer adducts. The phthalate esters plasticizers are commonly leached from plastic ware and were seen in ESI-MS spectrum as MH^+ ions. Low plasticizer laboratory ware or glass was used in all experiments. This hypothesis was tested by adding a cocktail of common plasticizers over the appropriate mass range. All flew well as MH^+ ions but did not form adducts with FKBP12; the hypothesis was rejected. The increased mass would be consistent with glycation of lysine. Glycation is a non-enzymatic process in which lysines react with reducing sugars at the ε amino group, this is thought to be catalysed by lysine next to a lysine in the primary sequence of a protein (Johansen *et al.*, 2006). The NetGlycate 1.0 Server (http://www.cbs.dtu.dk) predicts a score of 0.958 for K35 in FKBP12 (a number between **-1** and **1**; when the score is above **0** the residue is a predicted glycation site). K34 is contiguous to K35 in the primary sequence of FKBP12.



(A) ESI-MS (B) MALDI

Figure 2-7 Mass spectra of FKBP12.

(A) The transformed ESI-MS spectrum of 12.5 μM FKBP12 in 10 mM ammonium acetate, 10 % methanol (v/v) at pH 6.8. The dominant signal was at 13281.±4 Da. The spectrum was background corrected, smoothed and transformed using MassLynx V4.1 (Waters). (B) The MALDI mass spectrum of FKBP12 recorded on a sinapinic acid matrix. The major monomeric peak corresponds to 13286 Da with a small subsidiary signal at 13469 Da.

ESI-MS provided a convenient method of checking both the purity of FKBP12 and the presence of denatured protein. Low molecular weight breakdown products have been observed on SDS-PAGE gels over 8 weeks storage at 0 ^{o}C. These were easily picked up on an ESI-MS spectrum.

Globular proteins exhibit a narrow range of charge-states; 13+ to 7+ are characteristic of the six histidine tagged construct of FKBP12. Denatured protein exhibits a spectrum with a wider charge-state distribution with more highly charged ions indicative of basic residues exposed and available for protonation in the unfolded state. Figure 2-8 shows the effect of pH on the charge-state distribution of FKBP12. At pH 2.5, FKBP12 unfolds.

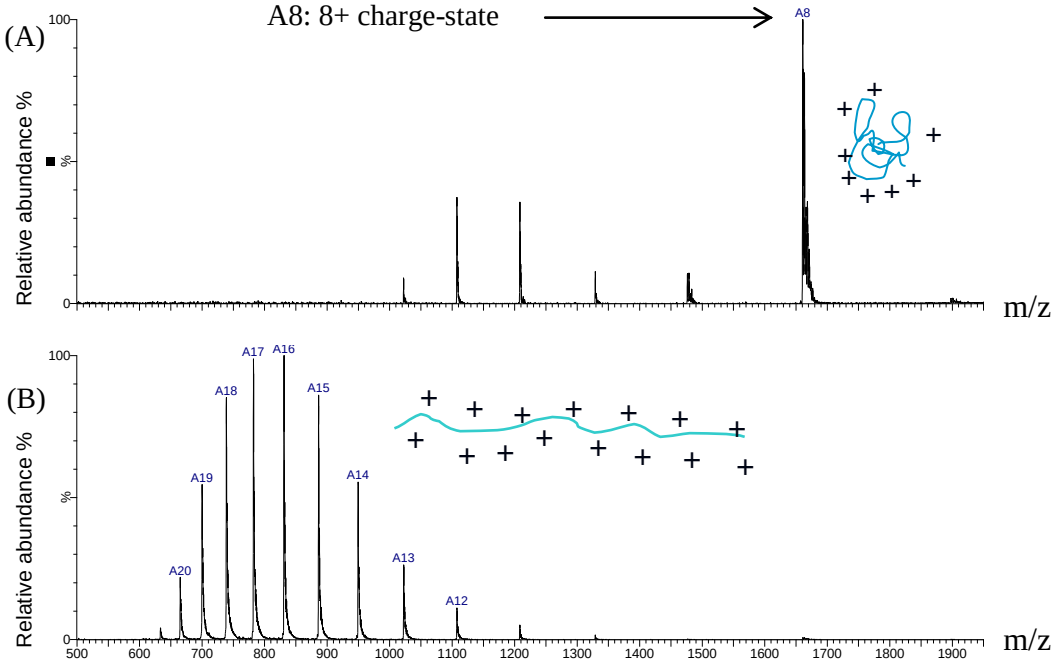


Figure 2-8 Effect of pH on the charge-state distribution of FKBP12.

(A) Representative ESI-MS spectrum of 12.5 µM FKBP12 in 20 mM ammonium acetate, 10 % methanol (v/v) at pH 6.8. Inset is a schematic of FKBP12 at pH 6.8. (B) ESI-MS spectrum of 12.5 µM FKBP12 in 20 mM ammonium acetate, 10 % (v/v) methanol at pH 2.5. Inset is a schematic of FKBP12 at pH 2.5 showing loss of tertiary structure. Species are labelled with charge-state; for example A8 is the 8+ charge-state.

2.3.4 **Protein characterisation by dynamic light scattering**

Dynamic light scattering (DLS) experiments were performed on dilute aqueous solutions of FKBP12 in 10 mM ammonium acetate, pH 6.8 and PBS at pH 7.5 to examine the quaternary structure and possible aggregation of FKBP12 in dilute solutions of common buffers (Horizon DLS, Viscotek).

DLS is a useful indicator of the oligomeric state of a protein and the detection of aggregate species in solution. Molecular diffusion is function of particle size. The effective hydrodynamic radius of a protein species can be estimated from the intensity of fluctuations in laser light scattered by molecules undergoing Brownian motion in solution. Correlation between intensity fluctuations and diffusion rates for molecules of different sizes can be made because intensity fluctuations are not random. Intensity functions may be modelled and correlated with the observed intensities (Wyatt, 1993). The scattering intensity of a molecule is proportional to the cube of its molecular mass and the molecular radius to the sixth power. DLS is therefore a sensitive tool to pick up changes in molecular radius of the species scattering the light, and hence the oligomeric state of the protein. Large species such as protein aggregates or unwanted particulate matter are easily detected.

Figure 2-9(A) shows an intensity distribution for an aqueous solution of FKBP12 in two common experimental buffers. The dominant species seen was a narrow peak centred between 2.4 and 2.6 nm hydrodynamic radii. The small peaks representative of larger particles are disproportionably abundant due to the non-linear relationship between scattering intensity and size ($I \alpha Radius^6$). When the data is presented as a mass distribution the radius of the dominant species are better represented as the mass contribution of the larger species are negligible, Figure 2-9(B). A radius of 2.5 nM is consistent with a monomer of FKBP12 at low concentration in ammonium acetate and phosphate buffers. These results were in agreement with gel filtration.

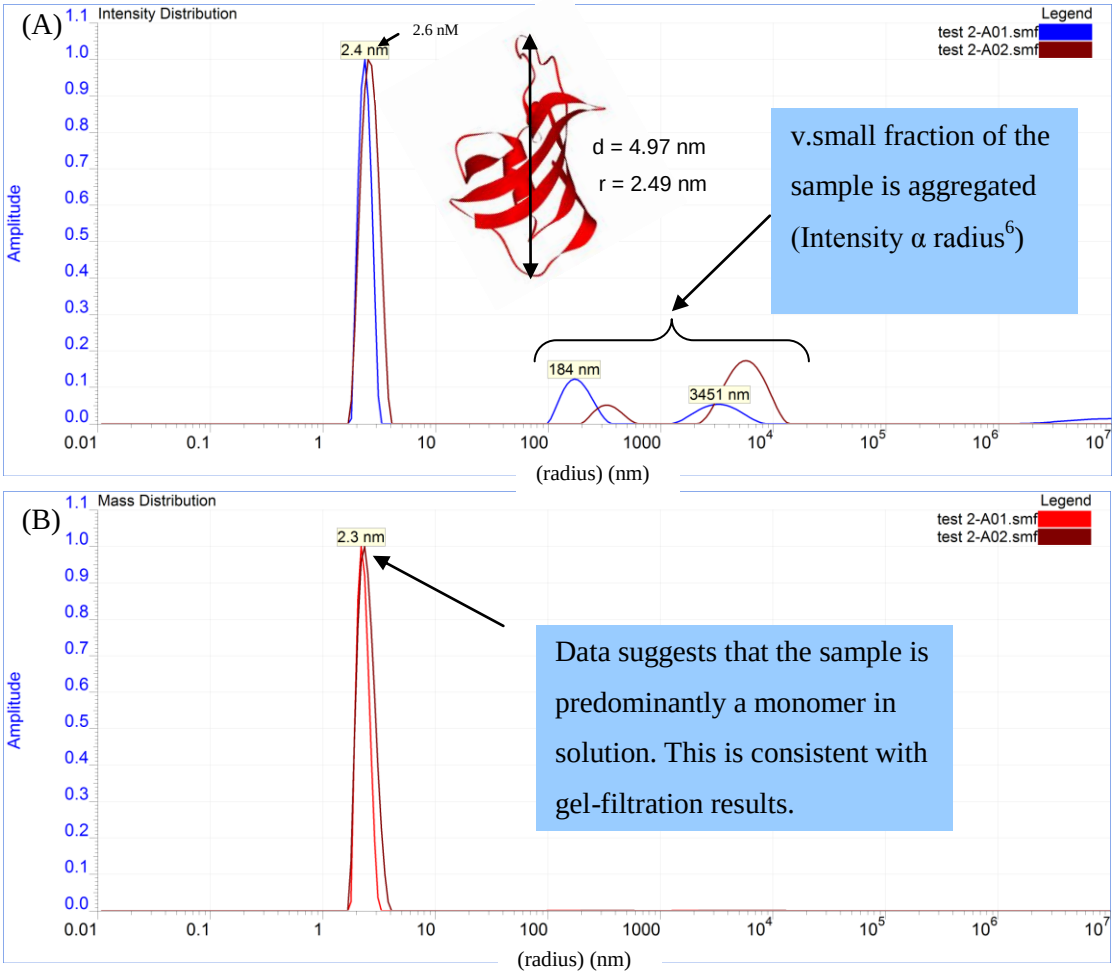


Figure 2-9 Dynamic Light scattering analysis of dilute aqueous solutions of FKBP12

A01 45 µM FKBP12, 50 mM ammonium acetate, pH 6.8; **A02**: 19µM FKBP12, PBS, pH 7.4, 1 mM DTT and 0.5 mM EDTA (A) Intensity distribution histogram, the x-axis shows hydrodynamic radius (nm) Inset shows a cartoon representation of FKBP with the molecular diameter labelled (PyMol V1.02, DeLano Scientific). (B) Mass distribution histogram, the x-axis shows hydrodynamic radius (nm). As the intensity of scattering from a particle is not directly proportional to size (Intensity α Radius^6) large particles of aggregated protein appear disproportionately abundant when viewed as an intensity distribution (A). If the data is transformed to a mass distribution as in Figure 2-9 (B) the distribution is a better representation of the relative proportions of the different species in solution.

2.3.5 **Determination of protein activity**

The FKBP12 construct used in this study to produce recombinant protein has been shown have ≥ 90 % activity for rapamycin when compared to wild type protein (Wear *et al.*, 2007b).

A Peptidyl-Prolyl Isomerase (PPIase) assay was carried out to determine the activity of batches of a purified FKBP12. The assay was performed as described in 6.3.1, following the method of Kofron *et al.* 1991.

3 Chapter 3 ESI-MS in drug discovery

3.1 Introduction

Electrospray ionisation mass spectrometry (ESI-MS) is an established tool for the study of non-covalent protein-ligand interactions in drug discovery (Hofstadler and Sannes-Lowery, 2006). It is a soft ionisation technique with the capacity to retain both tertiary structure and intact complexes (Ganen *et al.*, 1991; Loo, 1997). So called "direct detection methods" use syringe pump infusion of an analyte into the mass spectrometer and measure the relative abundance of different protein species. It is possible to obtain the mass of a ligand, the mass of the protein, binding stoichiometry and absolute equilibrium dissociation constants (K_d) for non-covalent interactions (Tjernberg *et al.*, 2004).

3.2 An experimental MS study of protein-ligand interactions

This study aimed to answer a number of key questions in the study of protein-ligand interactions by ESI-MS. Two questions are addressed in this chapter:

- What are the optimum conditions for observing protein-ligand interaction in MS for FKBP12?

- How do rank orders of affinity and K_d compare with data from experiments carried out in solution?

The binding of FKBP12 to FK506 and rapamycin were early examples of ESI-MS being used to examine a protein-ligand complex (Ganen *et al.*, 1991). The high molecular weight (>800 Da), natural product immunosuppressants FK506 (0.4 nM) and rapamycin (0.2 nM) have high affinity for FKBP12 (Bierer *et al.*, 1990). The non-covalent interaction profiles for FK506 and rapamycin are complex and include multiple hydrogen bonds and hydrophobic contacts (Van Duyne *et al.*, 1993). These ligands are therefore not particularly representative of the lower molecular weight compounds that are commonly selected at the early stage of the drug discovery process. It is thought that ESI-MS preferentially preserves complexes with a predominance of electrostatic interactions. It has been suggested that the influence of the hydrophobicity is less important in the gas phase due to the significant role for solvent (Robinson *et al.*, 1996; Daniel *et al.*, 2002). These observations are perhaps at odds with both the potential importance of hydrophobicity in contributing to a favourable Gibbs free

energy of binding between a target and potential drug. The active site of FKBP12 consists of a shallow hydrophobic pocket; it is therefore desirable to be able to detect relatively low affinity ligands where hydrophobic effects form a significant contribution to binding. In the context of screening collections of compounds against FKBP12, it is important to understand how ratios of complexed *versus* uncomplexed protein in the gas phase, compare with those in the solution for ligands with a range of affinity, hydrophobicity and mass. The applicability of the technique for diverse protein targets and the correlation of ESI-MS K_d with data derived from ITC or biochemical studies are still under much debate.

3.3 Principles of electrospray ionisation of protein molecules

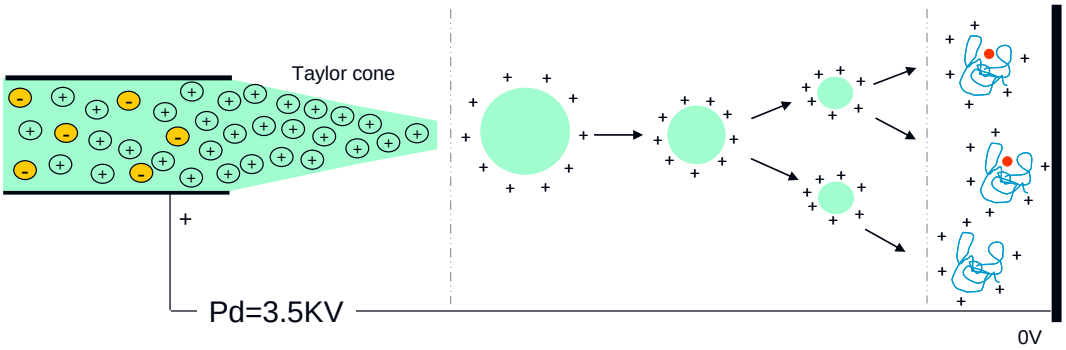


Figure 3-1 Schematic diagram of ion production during electrospray ionisation.

Charged drops are formed at the capillary tip and accelerate in an electric field. Drops are desolvated with the aid of a heated desolvation gas (N_2) and undergo repeated fission until a charged macromolecular ion formed.

In direct infusion ESI-MS run in the positive ion mode, the analyte is injected into a stainless steel capillary by means of a syringe pump. Basic residues on the surface of the globular protein become charged through reaction with NH^{4+} ions present in the analyte solution (Peschke *et al.*, 2002). In this study of FKBP12, ESI-MS was carried out at pH 6.8; at this pH the protein retains its folded conformation. Solution pH is a compromise between residue ionisation and maintaining the protein in the higher order structure. Denatured protein (pH 2.5) flies very well but is only useful in the accurate determination of protein mass. It is obviously desirable to test ligands at a physiologically relevant pH. The theoretical pI of FKBP12 is estimated to be 8.79 by the ExPASy ProtParam server (www.expasy.org/tools/protparam.html, June 2008, (Gasteiger *et al.*, 2005)).

Under the action of a strong electric field in the region of the capillary, cations in solution experience a force and undergo electrophoretic movement in the direction of the grounded plate (Figure 3-1). Surface tension counteracts the force experienced by the ions and a Taylor

cone of analyte, rich in positively charged ions, is formed at the capillary tip. Positively charged droplets break away from the Taylor cone and accelerate towards the plate. Evaporation of solvent from the charged drops is aided by collisions with a sheath of heated N_2 gas, the desolvation gas, Figure 3-2.

The explanation for the formation of desolvated protein ions has been much debated. The consensus is the charge residue model (CRM) better explains the production of macromolecules in electrospray and the ion evaporation model (IEM) best describes the production of small ions (Fernandez de la Mora, 2000). The CRM proposes that drops contract by the evaporation of solvent until they reach their Rayleigh stability limit (Rayleigh, 1882) and undergo Coulombic fission (Dole *et al.*, 1968). Repeated fission leads to droplets containing a single analyte molecule. The analyte molecules retain the droplet charge and become gas phase ions that subsequently enter the mass analyzer. The IEM model (Iribarne and Thomson, 1976) proposes that single solvated analyte ions are desorbed from the drop carrying some of the charge; thus reducing the Coulombic repulsion of the drop.

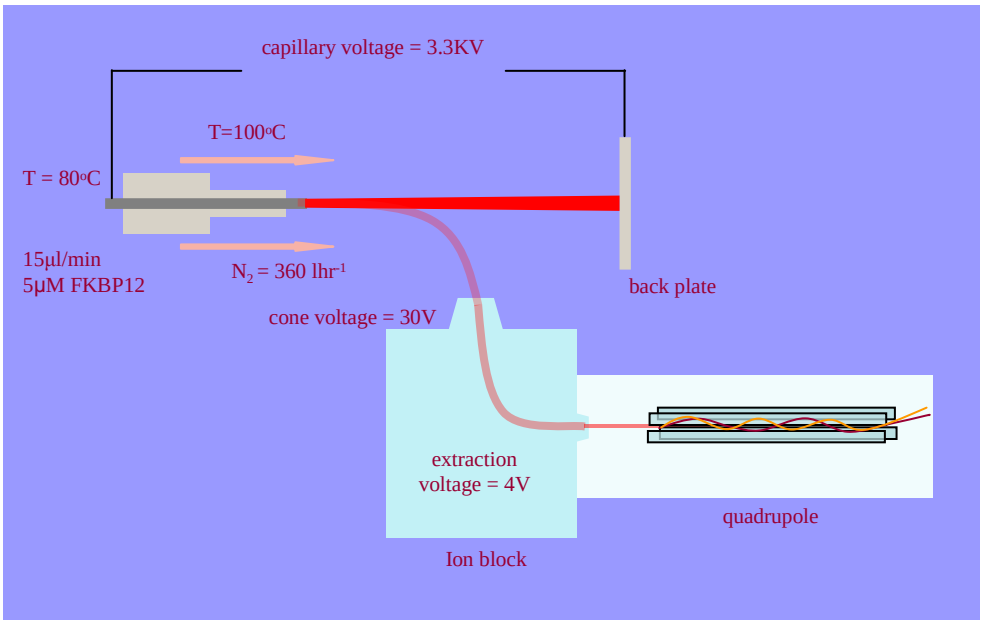


Figure 3-2 Schematic diagram Micromass ZMD under standard operating conditions.

In the Micromass ZMD spectrometer ions of the appropriate mass to charge range are taken through a "Z-bend" to the single quadrupole mass analyzer by a series of orthogonal electric fields. The diagram represents the stepped pressure reduction in the system by intensity of blue background. Un-desolvated charged drops do not enter the mass analyzer and hit the back plate.

3.4 The ESI-MS spectrum

A distinguishing feature of ESI-MS is that a single point experiment provides an electrospray spectrum with multiple signals. Each signal envelope results from a related population within the sample (Figure 3-3). The electrospray spectra of a protein, from a solution under conditions that retain tertiary structure, exhibit a characteristic series of multiply charged positive ions. The charge on the ion is referred to as the charge-state of the protein. Each charge-state represents a population with a defined mass to charge ratio. A charge-state spectrum with a narrow range of charge-states is indicative of the protein retaining its folded conformation (Peschke *et al.*, 2004b). When proteins are folded only basic residues on the surface of a globular protein are exposed for protonation.

In the analysis of small molecules by mass spectrometry, the deliberate fragmentation of a molecule often aids structural elucidation. The consideration of the instrument induced dissociation of a complex and the understanding that can be gained from that, if any, is a central theme that links many of the experiments in this study. The main instrument parameters that influence the dissociation of a protein-ligand complex are as follows:

- Source block temperature
- Desolvation gas temperature
- Capillary voltage
- Cone voltage

Results for optimising retention of the protein-ligand complex are presented in the following sections of this chapter.

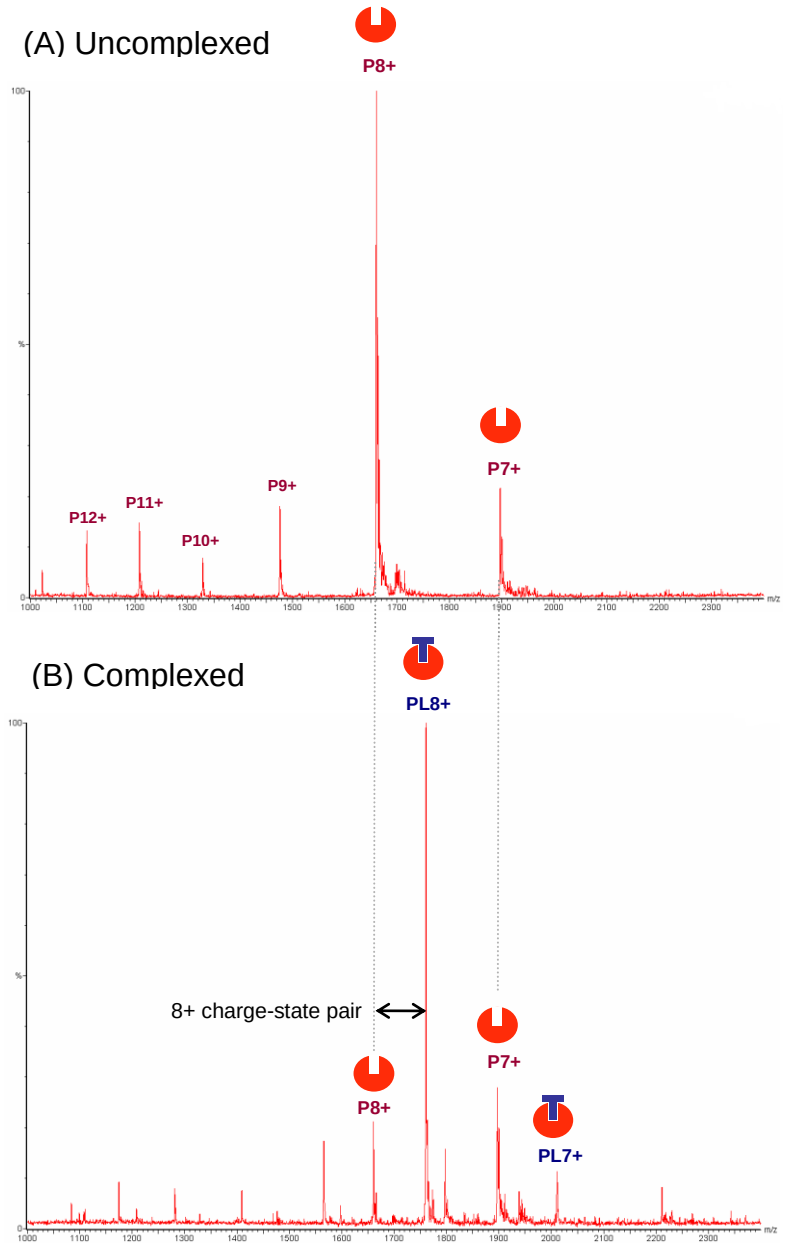


Figure 3-3 ESI-MS spectrum for FKBP12 and FKBP12 in complex with the immunosuppressant FK506.

(A) The electrospray spectrum of uncomplexed FKBP12. 10 µM FKBP12 in 10 mM ammonium acetate, 10 % methanol (v/v) at pH 6.8 recorded between 1000 - 2500 m/z. The charge-states 12+ to 7+ are labelled P12+ to P7+, the 8+ charge-state dominates the spectrum. (B) The electrospray spectrum showing the majority of the protein in complex with FK506. 10 µM FKBP12, 7 µM FK506 in 10 mM ammonium acetate, 10 % methanol at pH 6.8 recorded between 1000 - 2500 m/z. The 8+ and 7+ charge-states for FKBP12 are labelled P8 + to P7+; the 8+ and 7+ charge-states for the FKBP12-FK506 complex are labelled PL8+ and PL7+. Spectra were an average of 20 scans and have been background corrected and smoothed using a Savisky-Golay smoothing algorithm as part of MassLynx MS software package V4.1 (Waters).

3.5 Influence of instrument parameters on preservation of complex

3.5.1 Default instrument parameters

Parameter optimisation in ESI-MS is an iterative process (Figure 3-4). This is mainly because the conditions required to preserve the complex are mutually exclusive with a good signal to noise ratio. The specific influence of each instrument parameter is dealt with separately in the following sections, although the parameter set should be considered as a whole.

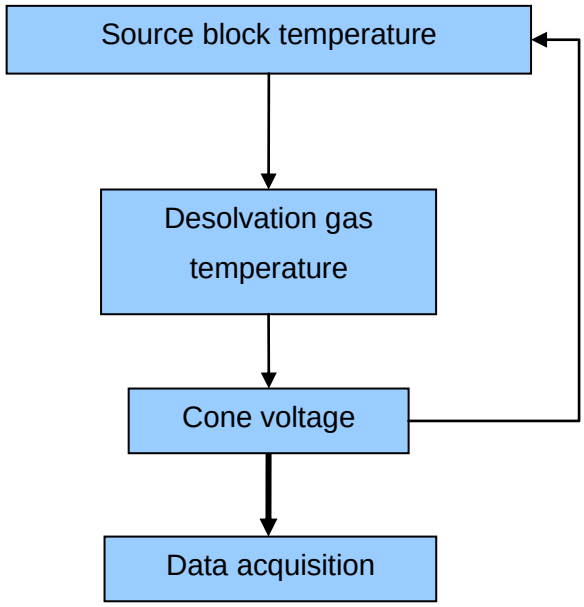


Figure 3-4 Parameter optimisation.
Parameter optimisation is an iterative process.

Electrospray mass spectrometry was carried out on three single quadrupole instruments, all employed atmospheric electrospray ionisation. Two were Z-spray devices; the Micromass ZMD and ZQ (Waters) and the third an ion-trap; the LCQ^{TM}DECA (ThermoQuest). FKBP12 was prepared and characterised as described in Chapter 2 and exchanged into 10-20 mM ammonium acetate, pH 6.8 at concentrations between 5 μM - 20 μM using pre-packed PD-10 rapid desalt columns (GE Healthcare). 10 % methanol (v/v) (HPLC grade, BDH) was added and thoroughly mixed by pipette prior to direct infusion. Methanol aided the desolvation process. Protein was denatured, by the addition of 2 % formic acid (v/v) to obtain a pH close to 2.5, and flown under the same conditions. All reagents were of the highest grade available.

Standard operating conditions for ligand binding studies using the Micromass ZMD instrument were as follows: nitrogen flow rate 360 lhr^{-1}; analyte flow rate 15 $\mu l \cdot min^{-1}$; capillary voltage 3.3 KV; cone voltage 15 – 30 V; source block temperature 80 ^{o}C and desolvation gas temperature 100 ^{o}C. Spectra were acquired over at least 20 scans (5s per scan), averaged, background corrected and smoothed using the Mass Lynx MS software package, V4.1 (Waters). Conditions for the Micromass ZQ were very similar to those of the ZMD (analyte flow rate 10 $\mu l \cdot min^{-1}$). The flow rates in the Z-spray devices are influenced by instrument dimensions and geometry.

Standard operating conditions for the LCQ^{TM}DECA were as follows: nitrogen flow rate 100 units, analyte flow rate 6 $\mu l \cdot min^{-1}$, capillary voltage 3.5 KV, cone voltage 25 V and capillary temperature 50 ^{o}C. Spectrum were averaged over 300 scans (2s per scan) and processed using the Bioworks Browser V 3.0 software package (ThermoScientific).

Any changes to the standard operating conditions are detailed in the text.

3.5.2 Effect of source block temperature on the fraction of complex preserved in ESI-MS

The van't Hoff equation (equation 3) describes how the ligand K_d shows a dependence on temperature (Chaires, 1997). It would therefore seem desirable to perform electrospray screening experiments close to a physiologically relevant temperature. However, this is was not possible for the instruments employed in this study. Signal strength in electrospray is a function of temperature. ESI-MS analysis of small organic molecules is routinely carried out above 140 ^{o}C to increase the signal to noise ratio and give good desolvation. These temperatures are obviously incompatible with preserving the integrity of the protein. In addition, measuring absolute temperatures in electrospray is difficult. The source block, containing the electrospray capillary, is held at a user defined temperature. However, the analyte does not necessarily reach the source block temperature during rapid capillary transit. In addition evaporation of the solvent from charged drops reduces analyte temperature. The internal energy distribution of the analyte molecules is strongly influenced by the expansion of the ions entering the first low pressure sector of the instrument. The pressure differential between the spray region and the mass analyzer sector of the instrument causes a supersonic expansion that has the effect of reducing molecular internal energy (Gabelica and De, 2005). This "cooling" effect is opposed by the acceleration of the ions in the medium pressure sectors of the instrument and is a function of the cone voltage. The final temperature at the end of desolvation is almost certainly lower than the measured source block temperature. Van't Hoff analysis using source block temperature as a measure of T did not give meaningful results for known protein-ligand systems (data not shown).

It can be seen in Figure 3-5 (A and B) that the fraction of protein seen in complex with ligand reduced as source block temperature increased. The intensity signals corresponding to FKBP12 in complex with FK506 (PL) and uncomplexed FKBP12 (P) were seen to diverge above 90 ^{o}C. Source block temperature was set at a default value of 80 ^{o}C; this was felt to be the lowest temperature at which signal to noise ratio was acceptable. Ammonium ions have a low energy of dissociation; temperatures above 25 ^{o}C should be sufficient for thermal dissociation (Peschke *et al.*, 2004a).

The relatively high capillary temperature of the Micromass ZMD and ZQ (Waters) electrospray source was probably a limiting factor in detecting low affinity ligands. Experiments carried out on the LCQ^{TM}DECA (ThermoQuest) allowed lower capillary

temperatures to be used (50-70 ^{o}C). This instrument had the additional disadvantage of slow acquisition times due to the increased noise in the electrospray spectrum. The increased noise is in part due to the direct path from capillary to ion-trap; solvated macromolecular ions and buffer ions directly enter the mass measuring sector of the instrument due to the absence of the "Z-bend" (ZMD and ZQ instruments). Cone voltage induced dissociation was less straightforward on this instrument (described in section 3.5.4). Increasing cone voltage increased all species, not just the species of interest, leading to both signal to noise problems and collision induced dissociation (data not presented). These factors outweighed the slightly lower capillary temperatures possible on this instrument.

3.5.3 Effect of desolvation gas temperature on the fraction of complex preserved in ESI-MS

Increasing the desolvation gas temperature between 40 ^{o}C and 120 ^{o}C had little effect on the proportion of FKBP12 seen in complex with the ligands tested, Figure 3-5 (C and D). However, it did affect the degree of desolvation of the protein. This has implications for measurement of the protein and complexed species during screening experiments. ESI-MS spectra from an analyte that had undergone good desolvation showed charge-state envelopes with a small adduct tail. The adduct tail is the result of incomplete dissociation of non-specifically bound low molecular weight species such as water. This means that the maximum peak intensity for a charge-state envelope can be used to calculate the fractional contribution of the species. Good desolvation leads to little overlap between the adduct tail of the uncomplexed protein envelope and the peak corresponding to a protein-ligand complex. Minimising overlap removes the need for mathematical deconvolution. An analysis of the influence of desolvation on the estimation of K_d was carried out on the LCQ^{TM}DECA (ThermoQuest) and is described in section 5.7. The desolvation gas temperature was set at a default value of 100 ^{o}C on the Micromass ZMD and ZQ (Waters). The desolvation gas temperature is not adjustable on the LCQ^{TM}DECA (ThermoQuest).

It is also worth considering how possible changes to the hydration of a protein affect its structure. It is not possible to determine structure directly in ESI-MS; however investigation of non-covalent interactions by this method makes the assumption that there are no radical changes to structure on desolvation. Computational simulations of desolvation suggest that the RMSD of a desolvated structure compared to a reference structure increases with increasing desolvation (Patriksson *et al.*, 2007). It would seem reasonable not to over

desolvate a protein from the perspective of both structural change and possible fragmentation.

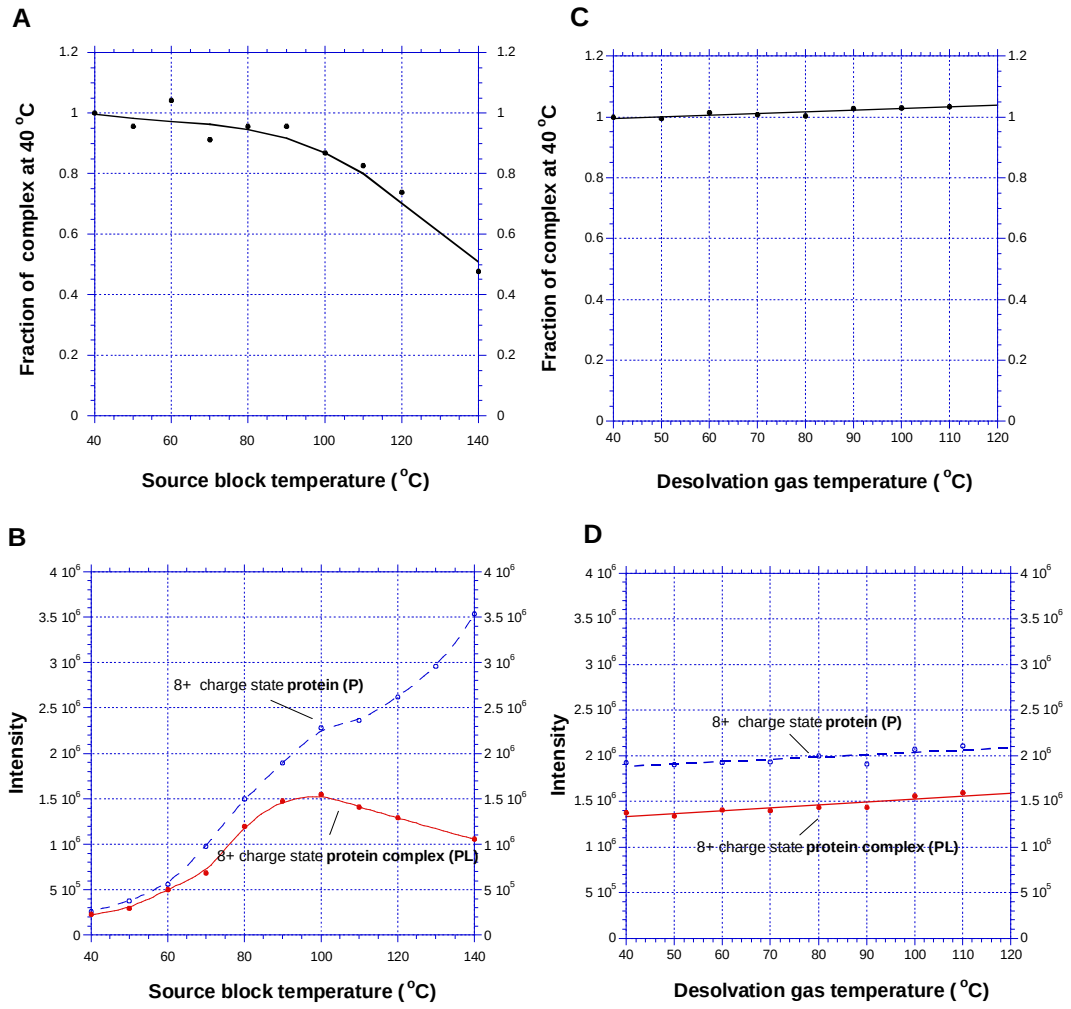


Figure 3-5 Fraction of protein complex preserved in electrospray is a function of source block temperature.

(A) Shows the fraction of a FKBP12-FK506 complex preserved as a function of source block temperature. Increasing capillary temperature above 90 °C lead to a rapid decrease in the proportion of complexed protein seen in the electrospray spectrum. (B) Shows the absolute signal intensities for uncomplexed FKBP12 and FKBP12 in complex with FK506 as a function of source block temperature. (C) Desolvation gas temperature had little effect on the fraction of complex observed between 40 and 110 °C. (D) Shows the absolute signal intensities for uncomplexed FKBP12 and FKBP12 in complex with FK506 as a function of desolvation gas temperature. 5.5 µM FKBP12, 2 µM FK506 in 10 mM ammonium acetate, 10 % methanol (v/v) at pH 6.8. Data is an average of 20 scans acquired on a Micromass ZMD instrument (Waters). Curves have been added to aid visualisation.

3.5.4 **Effect of cone voltage on fraction of complex preserved in ESI-MS**

Electrospray uses a source at atmospheric pressure; pressure is therefore reduced in stages as ions are drawn through the instrument to the mass analyzer. Ions pass from one region to another through a series of orifices or cones. The cone in an ESI-MS instrument can be set at a user defined potential to control the flow of ions through the instrument. Setting the cone voltage to an appropriate value is a very important part of parameter optimisation. The proportion of the protein seen to fly in complex with ligand is strongly correlated with cone voltage, Figure 3-6. Raising the cone voltage increases the speed of the macromolecular ions. An increase in speed increases the frequency and energy of molecular collisions and can result in dissociation of non-covalently bound complexes (Touboul *et al.*, 2008b). Deliberate dissociation by this method has been referred to as pseudo MS/MS (Yin *et al.*, 2008). The fraction of the protein that dissociates at a given cone voltage is a function of the affinity of the ligand and the nature of the molecular interactions (Rogniaux *et al.*, 1999)

Preserving non-covalent complexes of FKBP12 and the immunosuppressants FK506 and rapamycin in ESI-MS was relatively straightforward at low cone voltages. However, the signal corresponding to the FKBP12-FK506 complex rapidly diminished above 35 V, the signal for un-complexed FKBP12 increased. Section 4.3 discusses differences in cone voltages dissociation profiles for a range of ligands of FKBP12. Profiles are discussed in the context of the character of the non-covalent interactions between protein and ligand.

It is interesting to note that signal for the charge-state envelope at 10 V was more heavily adducted than charge-state envelopes seen at higher voltages. Figure 3-6 illustrates the reduction in the charge-state adduct tail as voltage increased and is consistent with higher cone voltages giving a cleaner signal with fewer non-specifically bound species. The variability in non-specifically bound adducts is investigated in section 5.7. The default value for cone voltage was set to 15 V for the Micromass ZMD and ZQ (Waters) and 25 V for the LCQ^{TM}DECA (ThermoQuest). This formed a starting point that was adjusted in individual experiments (values are detailed in the text).

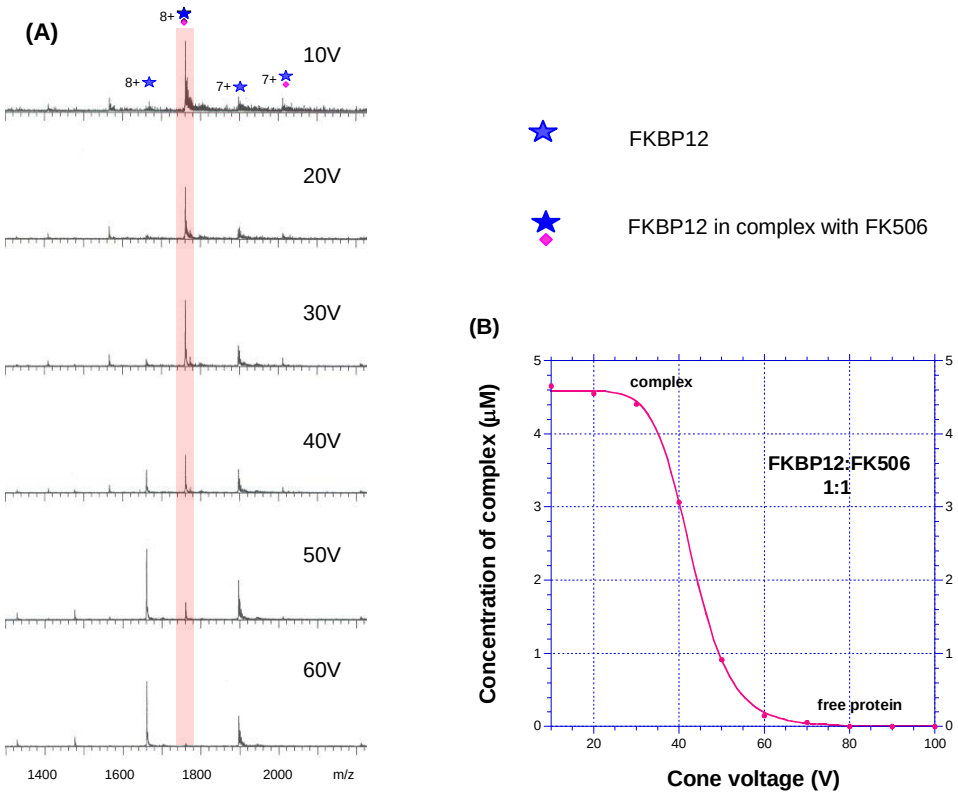


Figure 3-6 Fraction of non-covalently ligand bound to protein is a function of cone voltage.

Cone voltage had a significant influence on the fraction of complex preserved in electrospray. High cone voltage caused collision induced dissociation of the complex. (A) Raw spectra acquired between 10 and 60 V, data normalized for the largest peak in the spectrum. The pink highlight corresponds to FKBP12 flying in complex with FK506. The free protein signal is annotated with a blue star and the complex with a blue star and pink diamond. (B) The concentration of complex seen as a function of cone voltage calculated from peak intensities of the 8+ charge pair (section 3.6.3). 5.5 µM FKBP12, 5.5 µM FK506 in 10 mM ammonium acetate, 10 % methanol (v/v) at pH 6.8. Data is an average of 20 scans acquired on a Micromass ZMD instrument (Waters).

3.5.5 The influence of protein-ligand ratio on estimated affinity

The experiment to induce dissociation of the FKBP12-FK506 complex was repeated at two different ratios of protein to ligand. One at a 1.1 ratio of protein (FKBP12) to ligand (FK506) and another at 5:2 protein to ligand. The inverse relationship between the concentrations of complex (FKBP12-FK506) observed in ESI-MS and cone voltage is similar (inset diagram, Figure 3-7). The complex is very sensitive to cone voltage above 35 V for both analyte ratios. At a cone voltage of 60 V almost no complex is observed. The ratio of concentrations of complex observed in the two different analytes was used as a metric to compare cone voltage induced dissociation at different protein ligand ratios and given the symbol **W**.

$$\mathbf{W} = [PL]_{1:1\ ratio} / [PL]_{5:2\ ratio}$$

W was plotted in Figure 3-7 as a function of cone voltage (red line in Figure 3-7). The law of mass action predicts that **W** is 2.3; (black dashed line in Figure 3-7).

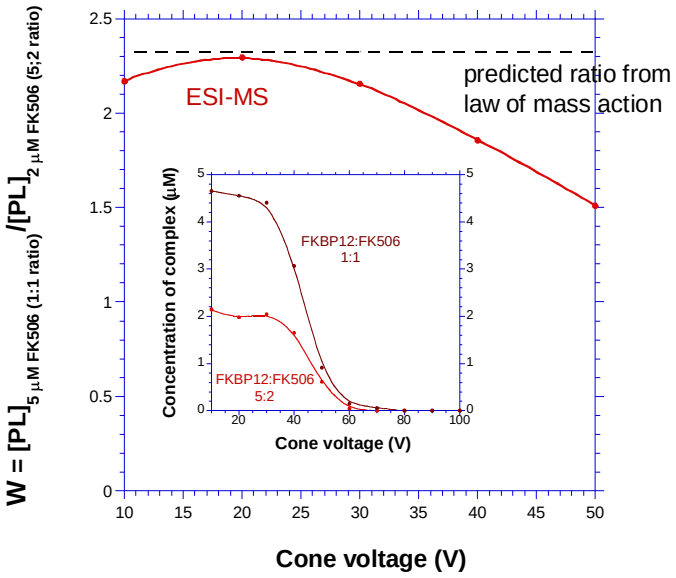


Figure 3-7 Does the analyte ratio influence complex dissociation in electrospray?

The black dashed line shows the theoretical solution phase ratio (**W = 2.3**) of FKBP12-FK506 complex for 1:1 and 5:2 ratios of protein to ligand in solution as calculated by the law of mass action. The experimental W is shown as red points. Inset diagram shows the concentration of complex as a function of cone voltage for the two protein ligand ratios. 5.5 µM FKBP12, 5.5 and 2 µM FK506 in 10 mM ammonium acetate pH 6.8; 10 % methanol (v/v). Source block temperature 80 ^{o}C, desolvation temperature 100 ^{o}C, cone voltage 10 to 50 V. Data is an average of 20 scans acquired on a Micromass ZMD instrument (Waters).

The actual ratio of complex measured in ESI-MS was compared to the theoretical solution phase ratio for each cone voltage calculated from the law of mass action (equation 1). The ESI-MS ratio was found to be very close to theoretical values for low voltage but diverged as cone voltage increased. This suggested that the electrospray ion intensities of P and PL are a better approximation to concentrations in solution if cone voltage is kept low. It is possible to rationalize these results in terms of the momentum of a given species and collision probability. More massive PL ions would have a lower mean speed (both in respect to their acceleration under the action of an electric field and the Maxwell-Boltzmann distribution). However, the PL ion would also have the larger collisional cross-sectional area and a slightly increased momentum. Do ions have a higher probability of dissociation if PL collides with PL rather than P hitting PL?

3.5.6 **Conclusions**

The six-histidine construct of FKBP12 flew well on the instruments tested and gave a narrow charge-state distribution dominated by the 8+ charge-state. The earliest ESI-MS studies on FKBP12, using His-tag free protein, found the charge-state distribution to be dominated by the 7+ charge-state (Ganen *et al.*, 1991). The difference in this study is likely to be due to the extra arginine present in the linker between the His-tag and native sequence, Figure 2-1 (Wear *et al.*, 2007b). A protein concentration between 5 and 10 μM was required to give a good signal to noise ratio; this was instrument dependent.

3.6 Estimating K_d from a titration experiment

3.6.1 Materials and Methods

For titration experiments, ligand was added at concentrations between 0.5 μM and 2 mM from stock solutions dissolved in the methanol fraction, no additional solvents were used. Ligands were stored at -80 ^{o}C in methanol, at concentrations between 25 mM and 100 mM (dependent on ligand solubility). The only exception to this rule were the experiments on hen egg white lysozyme (HEWL) where the addition of methanol was found to give a protein spectrum with poor desolvation. In the series of experiments with *N, N', N''*-Triacetylglucosamine (NAG3) the ligand was dissolved in deionised water. NAG3 was found to have very low solubility in methanol. Analyte solutions were carefully mixed by pipette and allowed to incubate at room temperature for at least 10 minutes.

The PL/P intensity ratio (R) was measured for the 8+ charge-state pair and K_d determined by the least squares fit of the data to equation 12 using the non-linear regression program KaleidaGraph version 4.03 (Synergy Software) as described in 3.6.3.

3.6.2 Estimating ligand affinity by the titration method

The calculation of the equilibrium dissociation constant from the relative signal intensities corresponding to complexed (PL) and uncomplexed protein (P) in a titration experiment is dependent on several assumptions. The principal assumption is that the ratio of ion intensities of PL (I_{PL}) and P (I_P) are proportional to those in solution. In addition, it is assumed that equilibrium relaxation times are long compared to solvent evaporation time. If the analyte concentration increases as solvent evaporates and the equilibrium shifted towards PL it is possible that the affinity of a ligand may be over estimated. The opposing effect of collision induced dissociation of weak non-covalent complexes would result in the affinity of the ligand being underestimated. A third assumption is that very weak non-specific interactions are not retained in the final desolvation step. This study found that complete desolvation is incompatible with conditions necessary to preserve non-covalent interactions between protein and ligand. Mass spectra peak envelopes inevitably show adduct tails under ligand screening conditions. Peschke *et al*. (2004a) examined the above assumptions in the context of the charge residue model. The researchers concluded that deviations from the above assumptions increase as the ligand approaches the mass of the protein. The ligands of FKBP12 examined in this study are all < 10 % of the protein mass.

3.6.3 Calculating ligand mass and the equilibrium dissociation constant

Figure 3-3 (A) shows the spectra for a sample of FKBP12 and Figure 3-3 (B) a similar spectrum after the addition of the high affinity ligand FK506. The spectrum shown in Figure 3-3 (B) can be viewed as two distinct charge-state distributions; one for the free protein and one for the protein in complex with FK506 (see peak annotation). For a charge-state of a given magnitude there are two peak envelopes. One represents the free protein and one the ligand in complex with the protein; this is referred to as a charge-state pair. Figure 3-8 illustrates the 8+ charge-state pair for FKBP12 and the FKBP12-FK506 complex. When viewing the full m/z range, spectrum for the complex can be thought of as a separate charge-state distribution (in addition to that of the free protein) transposed towards a larger m/z. It is possible to calculate the mass of a ligand bound to protein from any charge-state pair. For example, the mass of FK506 may be calculated from equation 4 using the 8+ charge pair.

$$m_{FK506} = 8((m/z)^{PL8+} - (m/z)^{P8+}) \qquad \textbf{Equation 4}$$

Where: m_{FK506} is the mass of FK506, $(m/z)^{PL8+}$ the mass to charge ratio of the FKBP12-FK506 complex and $(m/z)^{P8+}$ the mass to charge ratio of uncomplexed FKBP12.

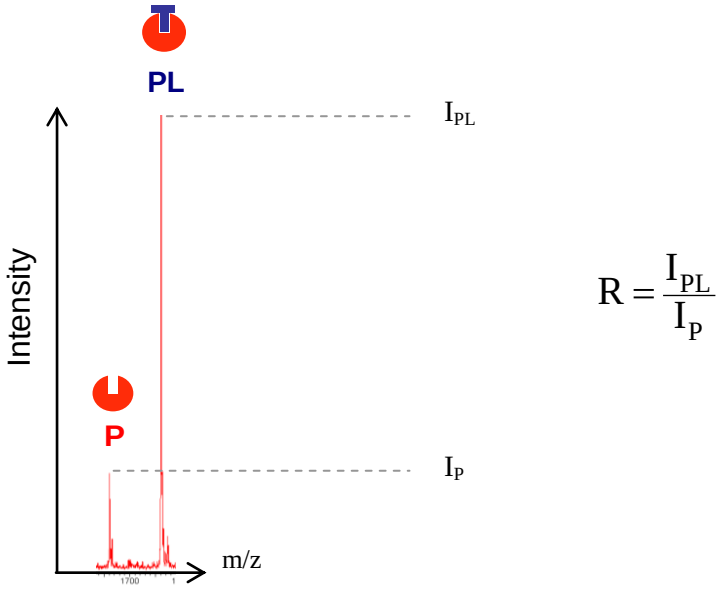


Figure 3-8 ESI-MS spectrum for the 8+ charge-state pair of FKBP12 free (P) and in complex with the immunosuppressant FK506 (PL).

10 µM FKBP12, 7 µM FK506 in 10 mM ammonium acetate, pH 6.8; 10 % methanol recorded between 1000 - 2500 m/z. Spectra were averaged over 20 scans, background corrected then smoothed using Mass Lynx MS software package V4.1 (Waters).

The relative affinity of ligand for a protein may be estimated from the ratio, R, of signal intensity for complexed (I_{PL}) and uncomplexed protein (I_P) in a charge-state pair. The equilibrium dissociation constant, K_d, can be calculated using a model based on the law of mass action (equation 1, section 1.8). The method used in this study was based on that of Tjernberg *et al.* (2004).

For the purpose of the calculation, the assumption made was that ratio of species in the gas phase is equal to the ratio in solution (Peschke *et al.*, 2004a).

$$R = \frac{I_{PL}}{I_P} = \frac{[PL]_{(aq)}}{[P]_{(aq)}} \qquad \textbf{Equation 5}$$

Where: R is the ratio of intensity of complexed to un-complexed protein in the gas phase measured by ESI-MS; $[PL]_{(aq)}$ is the concentration of complex and $[P]_{(aq)}$ is the concentration of free protein in solution at equilibrium.

The equilibrium between free protein and protein in complex with ligand is represented by equation 6. Square brackets have been omitted to aid clarity.

$$P + L \rightleftharpoons PL \qquad \textbf{Equation 6}$$

Where: P is the concentration of the free protein, L the concentration of the free ligand, PL the concentration of the protein-ligand complex.

It is known that the sum of concentrations of free protein, P and complexed protein, PL is equal to the total protein concentration, P_t; represented in equation 7. Equation 8 represents the similar relationship for the ligand. Where: L_t is the total concentration of the ligand.

$$P_t = P + PL \qquad \textbf{Equation 7}$$

$$L_t = L + PL \qquad \textbf{Equation 8}$$

Rearranging equation 5, and substituting into equation 7:

$$P_t = \frac{PL}{R} + PL$$

$$\Rightarrow PL = \frac{R \cdot P_t}{(1+R)} \qquad \textbf{Equation 9}$$

Equation 9 allows one to calculate PL for each point in the titration by measuring I_{PL} and I_P in ESI-MS and calculating R using equation 5 (each titration point employs a different L_t and P_t is fixed). The law of mass action is represented in equation 1 (section 1.8) Where: K_d is the apparent equilibrium dissociation constant.

$$K_d = \frac{P \cdot L}{PL}$$

Rearranging equations 7 and 8 and substituting into the law of mass action leads to equation 10.

$$K_d = \frac{(P_t - PL)(L_t - PL)}{PL} \qquad \textbf{Equation 10}$$

Expanding equation 10:

$$K_d \cdot PL = P_t \cdot L_t - P_t \cdot PL - L_t \cdot PL + PL^2$$

Rearranging equation 10 to give a quadratic leads to equation 11.

$$PL^2 - (P_t + L_t + K_d)PL + P_t \cdot L_t = 0 \qquad \textbf{Equation 11}$$

The solution of the quadratic equation leads to equation 12.

$$PL = \frac{(P_t + L_t + K_d) - \sqrt{(P_t + L_t + K_d)^2 - 4P_t \cdot L_t}}{2} \qquad \textbf{Equation 12}$$

The equilibrium dissociation constant, K_d, can be calculated by the best fit of ESI-MS titration data to equation 12 using a non-linear regression program.

It is of course possible to determine the mass of many of the different species in solution if they charge positively during electrospray. The dominant species in the spectrum is often the $[L + H]^+$ ion where L is the ligand. The intensity of this peak is dependent on the gas phase basicity of the ligand, the fraction of the total ligand in complex with the protein in solution and the degree of dissociation of the complex during electrospray. The sodium adduct of the ligand, $[L + Na]^+$ is also often seen. Sodium ions are largely removed from protein solutions by desalting columns. However, there can be sodium ions associated with the ligand.

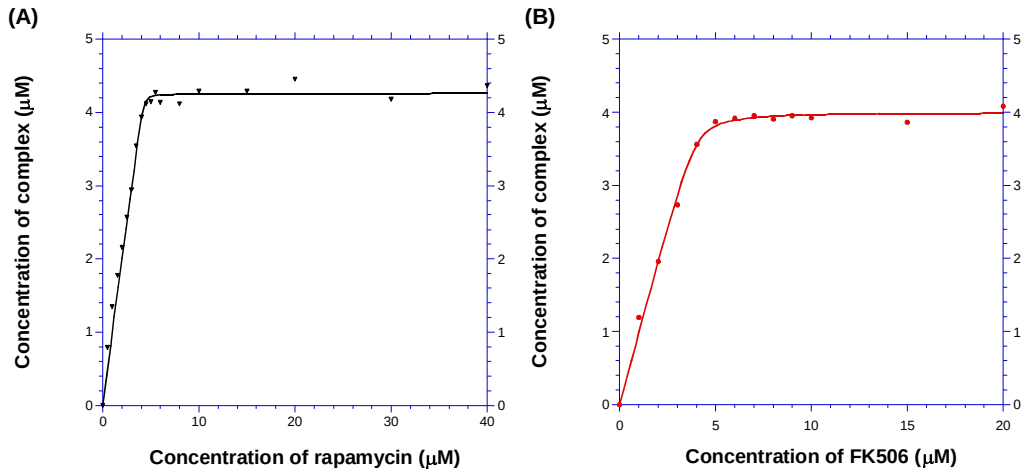
3.6.4 Correlation of affinity measured by ESI-MS titration and in solution

K_d values were estimated for the nanomolar affinity ligands of FKBP12 (rapamycin and FK506) and compared to those determined in the solution, Figure 3-9. The PL/P intensity ratio (R) was measured for the 8+ charge-state pair (the dominant protein charge-state) over a range of concentrations of inhibitor while fixing the concentration of protein. K_d was determined by fitting equation 12 to the data using the non-linear regression program KaleidaGraph (version 4.03, Synergy Software). The K_d apparent for rapamycin in complex with FKBP12 was very close to literature values. The K_d apparent for FK506 was underestimated by a factor of around 10. Figure 3-6 shows that at the experimental cone voltage necessary for a very good signal to noise ratio (30V) there was measurable collision induced dissociation; this is consistent with an underestimation of K_d apparent for FK506. The crystallographic evidence suggests that FK506 makes one fewer hydrogen bond with FKBP12 than rapamycin (2DG3.pdb, 1FKJ.pdb; described in Table 4-3). The extra hydrogen bond may account for the greater robustness of the FKBP12-rapamycin complex.

It is important to note that biophysical and biochemical affinity is not reported at a standard temperature in the literature. Temperature correction is possible when full thermodynamic data is available. However as previously discussed, accurate temperature measurements were difficult for the ESI-MS instruments used in this study.

The physical processes that govern the validity of the assumptions regularly employed in ESI-MS titration studies are complex; they are protein, ligand and instrument dependent. This is due to differences in temperature, spray drop size, desolvation time/efficiency and the nature of the non-covalent interactions between binding partners (Benkestock *et al.*, 2004). It

was therefore important to establish the sensitivity of a given instrument using a test set of inhibitors that have been characterised in solution.



(C)

K_d (nM)	*Ligand*	*Radio ligand displacement (25 °C)*	*Surface plasmon resonance (25 °C)*	*PPIase (4 °C)*	*ITC (25 °C)*	ESI-MS (>25 °C)
FKBP12	rapamycin	0.2[a]	1.67 ± 0.16[b]	7.8 ± 3.7[c]	1[d]	**0.18 ± 6.4**
FKBP12	FK506	0.4			0.5[d]	**59.6 ± 15.6**

a (Bierer *et al.*, 1990), b (Wear and Walkinshaw, 2007), c (Wear *et al.*, 2007b), d (Connelly and Thomson, 1992)

Figure 3-9 Measuring K_d by ESI-MS titration.

Typical experimental data from titration experiments with high affinity ligands. (A) 5 μM FKBP12, 0 to 40 μM rapamycin. (B) 5 μM FKBP12, 0 to 20 μM FK506. Analytes were dissolved in 10 mM ammonium acetate, pH 6.8: 10 % methanol (v/v). (C) A comparison of K_d from the literature with those measured by ESI-MS. Data is an average of 20 scans acquired on a Micromass ZMD instrument (Waters) under standard operating parameters (cone voltage 30 V). The solid lines are least squares fit to equation 12 using KaleidaGraph version 4.03 (Synergy Software).

3.7 Screening the inhibitor test set

3.7.1 Compound selection

One of the principal aims of this project was to explore medium throughput biophysical screening strategies for testing hits from virtual screening programs. To this end, it was necessary to compile a test set of compounds. Molecules were selected considering mass, structural diversity and affinity for FKBP12.

The inhibitor test set was used to examine the relative sensitivity of the ESI-MS instruments used in this study and to investigate the fraction of protein seen in complex in the gas phase compared to that seen in solution for a diverse series of ligands. All ligands have been reported to bind specifically to the active site of FKBP12 (Bierer *et al.*, 1990; Stebbins *et al.*, 2007; Wear *et al.*, 2007b). Early versions of the test set included compounds that were insufficiently soluble or active to be useful; these were removed. The final version is shown in Table 3-2 Ligand test set.. The ligands were detected as $[L + H]^+$ ions and as the major ionisable species under non-fragmenting conditions (positive ion mode, where L is the screening compound). Compounds SD1 to SD12 were selected from a high-throughput NMR screen of small molecules against FKBP12 performed by the San Diego Centre for Chemical Genomics. Affinity data was retrieved from the PubChem BioAssay database (SDCCG-A014-FKBP12-NMR, http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=608, September 2008, (Stebbins *et al.*, 2007). The data set is referred to as the San Diego data set in this thesis. Test set compounds were selected by choosing the most active purchasable member of a chemically similar cluster. Compounds were soluble to at least 25 mM in methanol and were stored at -80 ^{o}C.

Affinity	*Mass*	*No. of rings*	*No. of hydrogen bond donors*	*No. of hydrogen bond acceptors*	*MlogP*
0.2 nM - 686 µM	201 – 914 Da	2 - 3	0 - 4	3 - 14	0.44 – 2.05

Table 3-1 Physicochemical descriptor ranges for the ligand test set.

Affinity data from the literature (Bierer *et al.*, 1990; Stebbins *et al.*, 2007). Descriptors were calculated using the program Accord for Excel V6.1 (Accelrys Software Inc.). Note: the maximum number of hydrogen bond acceptors is 6 if rapamycin and FK506 are not included.

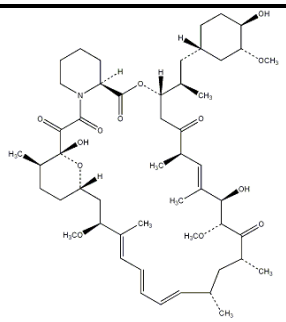
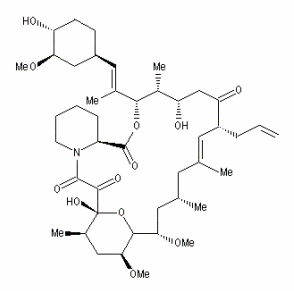
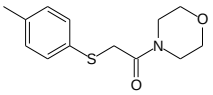
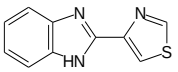
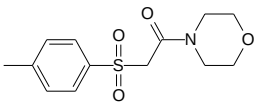
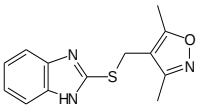
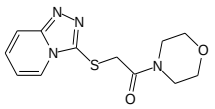
Compound	*K_d (μM)*	*MW (Da)*	*Structure*	*Reference*
rapamycin	0.0002	914.2		(Bierer *et al.*, 1990; Wear and Walkinshaw, 2007; Wear *et al.*, 2007b)
FK506	0.0004	804.0		(Bierer *et al.*, 1990)
SD1	9 (6.8) *	251.1		(Stebbins *et al.*, 2007)
SD12	56	201.0		(Stebbins *et al.*, 2007)
SD2	95	283.1		(Stebbins *et al.*, 2007)
SD3	157	259.1		(Stebbins *et al.*, 2007)
SD4	221	278.1		(Stebbins *et al.*, 2007)

Table 3-2 Ligand test set.

Structure, mass and affinity data for the FKBP12 test set. Ligands were selected for specificity to the active site of FKBP12, range of affinity and solubility in aqueous buffer. * Determined in this study by ITC (section 6.4.2). The table is continued overleaf.

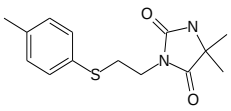
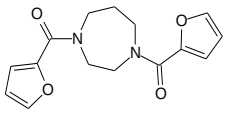
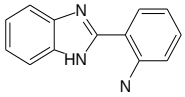
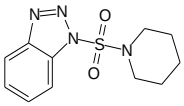
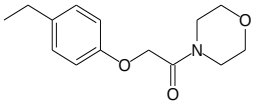
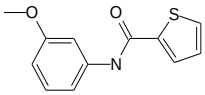
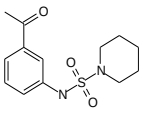
Comp.	K_d *(μM)*	*MW (Da)*	*Structure*	*Reference*
SD5	270	278.1		(Stebbins *et al.*, 2007)
SD6	277	288.1		(Stebbins *et al.*, 2007)
SD7	286	209.1		(Stebbins *et al.*, 2007)
SD8	385	266.1		(Stebbins *et al.*, 2007)
SD9	435	249.1		(Stebbins *et al.*, 2007)
SD10	440	233.1		(Stebbins *et al.*, 2007)
SD11	686	282.1		(Stebbins *et al.*, 2007)

Table 3-3 Ligand test set (continued)

Structure, mass and affinity data for the FKBP12 test set. Ligands were selected for specificity to the active site of FKBP12, range of affinity and solubility in aqueous buffer.

3.7.2 **The protein: ligand screening ratio**

We were interested in the suitability of ESI-MS for estimating the affinity of a ligand and creating a rank order of affinity. Some consideration was given to the ratio of protein to ligand most appropriate for an experiment at a single ligand concentration. To minimise uncertainties in the data it was desirable to measure I_{PL} and I_P when both signals had a good signal to noise ratio. Protein charge-state envelopes, under standard operating parameters, showed small adduct tails. We determined that an I_{PL}/I_P ratio, $R \geq 0.1$ provided a reasonable degree of certainty that the species was due to the protein ligand complex and not a random adduct. The m/z of a species is obviously important in deciding whether a peak corresponds to a protein complex or a non-specific adduct/noise. The MassLynx software was set up to label ion series. A screening ratio of 1:10 was selected; a protein concentration of 10 μM and ligand of 100 μM. This ratio was chosen as a compromise between detecting low affinity ligands and reducing the possibility of non-specific interactions (Sun *et al.*, 2007). Careful examination of the law of mass action (equation 1, section1.8) shows that at this ratio, a ligand with a K_d of 500 μM (single site) would result in 16.4 % of the protein being in complex with the ligand in solution ($R \sim 0.2$). Many screening compounds were not very soluble in the analyte solution, it was important to select a screening concentration at which the ligand dissolved.

3.7.3 **Suppression of the protein signal**

An additional consideration was the charging of screening compounds during electrospray, these frequently appeared as both $[L + H]^+$ and $[L + Na]^+$ ions in the spectrum. High concentrations of easily ionisable ligands dominated the ion current, suppressing the protein signal and leading to a very poor signal to noise ratio for the protein. This effect was particularly evident with the addition of DMSO to analyte solutions. A 1 % solution of DMSO is equivalent to a 141 mM solution. DMSO is routinely excluded from MS analytes or chromatographically separated prior to analysis. This study found a concentration of 0.25 % (35 mM) to be the maximum working value. It was also important to remember that DMSO had been shown to bind to FKBP12 with a K_d of 20 mM (Burkhard *et al.*, 2000). Thus adding DMSO to any screening experiment with FKBP12 leads to competition for the active site (this is discussed further in section 6.2).

3.7.4 **Rank orders of affinity from ESI- MS**

The desolvation or "cleaning-up" process is designed to remove the majority of water and buffer components non-specifically bound to the protein. However as discussed in 3.5.4, over-zealous desolvation greatly reduced the sensitivity for weak non-covalent interactions. Various cone voltages were tested for each instrument and are detailed in the text.

Compound	*K_d (μM)*	*MW (Da)*	*Fraction in complex in solution (calc. from K_d)*	*Micromass ZQ % FKBP12 in complex with ligand (8+)*	*Micromass ZQ Rank (cone=15 V)*	*LCQ^{TM}DECA Rank (cone=25 V)*
rapamycin	***0.002***	*914.2*	*99.6*	*98±4*	*1*	*1*
FK506	***0.004***	*804.0*	*99.3*	*83±1*	*2*	*2*
KB2_61D	**6.6**‡**/2.1**†	368.4	94.4	75±10	3	3
SD1	6.8‡	251.1	91.9	32±15	4	4
SD12	**56**	201.0	62.6	-	destabilizes protein	destabilizes protein
SD2	**95**	283.1	50.0	19±4	5	5
SD3	**157**	259.1	38.0	12±8	6	6
SD4	**221**	278.1	30.5	-	no complex	no complex
SD5	**270**	278.1	26.5	-	no complex	no complex

Table 3-4 Summary ESI-MS results for ligand test set.

Standard instrument parameters, cone voltage 15 V. Data in *italics* employed a screening ratio of 9.5 μM FKBP12, 10 μM rapamycin/FK506. Standard typeface: 10 μM FKBP12, 100 μM SD1, SD2, SD3, SD4, SD5. Analytes were dissolved in 20 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). Errors are standard deviations for n = 3. ‡ and † indicates a K_d measured in this study by ITC and PPIase respectively, section 6.3.1. Other values are quoted from the literature; see Table 3-2.

The nanomolar affinity ligands rapamycin and FK506 were seen to fly in complex with protein at well below standard screening ratios on all instruments (Figure 3-9). However, they were relatively insoluble in the analyte (10 μM FKBP12; 10 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). Rapamycin and FK506 were soluble to 10 fold higher concentration when methanol was included compared to a similar solution excluding methanol. Rapamycin, FK506, SD1, SD2 and SD3 were seen in complex with FKBP12 on the LCQ^{TM}DECA (cone voltage of 25 V) and the ZQ instruments (cone voltage 15 V). Figure 3-10 shows typical spectra for these ligands, blue diamonds mark ions corresponding to FKBP12 in complex with ligand. Results are summarized in Table 3-4. Rank orders were

in agreement with the literature for both instruments except for compound SD12. The spectrum for this ligand suggests that the tertiary structure of the protein had changed; the protein spectrum was characteristic of partial unfolding. Changes to the spectrum was time dependent (data not shown), there was no signal corresponding to FKBP12 in complex with ligand.

High resolution x-ray structures often show water mediating hydrogen bond bridges between protein and ligand (Helms and Wade, 1998; Taylor *et al.*, 2008). The water molecule may be thought of as an extension to the protein surface. If a water molecule is integral to ligand binding and is effectively trapped at the protein-ligand interface it should be possible to detect the mass of the water in MS. For example, if one water molecule were present the minimum mass of the signal corresponding to the complex would be [P + L + H_2O]. The minimum mass in a charge-state envelope were compared for the signal corresponding to the free protein and the protein in complex with the ligand (section 3.6.3). A peak was present in the spectrum corresponding to un-adducted protein plus the mass of ligand for all charge-states for the rapamycin, FK506, SD1, SD2 and SD3 complexes; no trapped waters were detected. X-ray structures do not suggest that hydrogen bonds between rapamycin and FK506 and FKBP12 are mediated by water. The ESI-MS data was consistent with the crystallographic data.

The fraction of the protein in complex with the ligand was calculated from the 8+ charge-state pair for data collected on the ZQ instrument (following the method described in section 3.6.3)

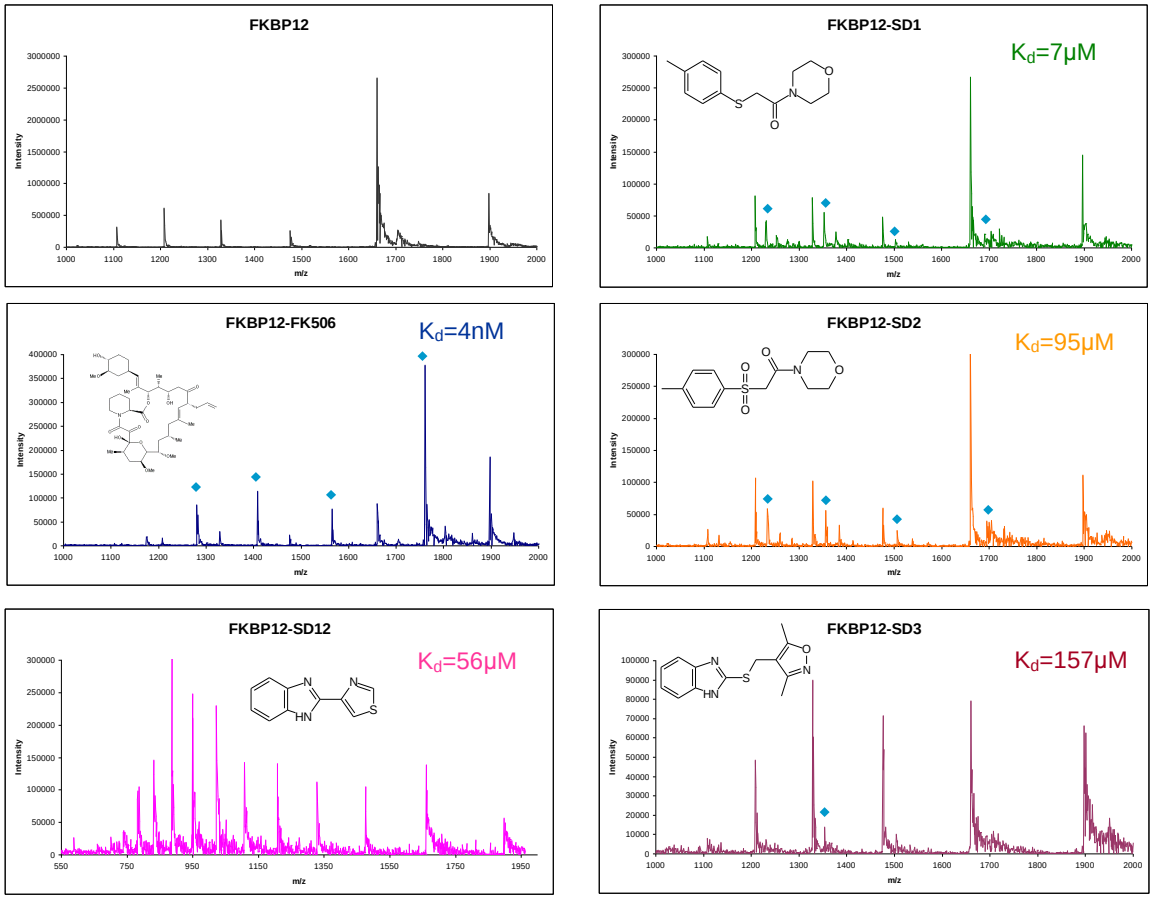


Figure 3-10 ESI-MS spectra for FKBP12 and ligands over a range of affinity.

Representative ESI-MS spectra for FKBP12 and a range of ligands, (1000 to 2000 m/z) ligands from the LCQ^{TM}DECA instrument. Ligand SD12 is shown between 550 and 2000 m/z to illustrate the destabilizing effect of the compound on the protein. A spectrum of free FKBP12 is shown for comparative purposes. 5 μM FKBP12, 100 μM ligand dissolved in 20 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). Blue diamonds mark peaks corresponding to FKBP12 in complex with ligand. K_d is quoted from the literature (Bierer *et al.*, 1990; Stebbins *et al.*, 2007; Wear *et al.*, 2007b).

3.7.5 Influence of charge-state on estimated affinity

It is interesting to note that there were subtle differences in the state distributions corresponding to the complex of some low affinity ligands with FKBP12. Free FKBP12 and FKBP12 in complex with the high affinity ligands rapamycin and FK506 have distributions dominated by the 8+ charge-state. SD1 and SD2 have distributions with a higher proportion of the complexed species flying in a charge-state > 8+. This phenomenon was not instrument dependent; similar spectra were obtained for both the ZQ and LCQ^{TM}DECA instruments. The effect was most marked for ligand SD1 on the LCQ^{TM}DECA, Figure 3-10. It is clear that changes to the charge-state distribution on ligand binding have implications for calculating PL from the peak intensities in a single charge-state pair. The 8+ charge-state was the dominant charge-state for free protein and protein complex in the majority of experiments and gave the best affinity discrimination between ligands in the test set at standard screening conditions.

3.7.6 Increasing the ratio of ligand to protein

SD1, SD2, SD6 and SD10 were titrated against 5 μM FKBP12 in a series of experiments on the Micromass ZMD. These ligands were selected for their range of affinity and solubility (0 to 1000 μM in analyte buffer). The cone voltage was set to 30 V, this was necessary to account for suppression of protein signal due to the high concentration of ligand. SD1 was not seen to fly in complex, SD2 was seen to fly in complex above 500 μM, SD6 above 250 μM and SD10 was never seen in complex. Titrations were also performed for the negative controls N-phenyl-naphthylamine (NPN), imidazole and adenosine monophosphate (AMP). These compounds were chosen as they are not known to bind to FKBP12 and charge differently in solution at pH 6.8. NPN and imidazole did not fly in complex with FKBP12 at the screening concentration. FKBP12 was seen to form a complex with AMP at a concentration above 500 μM.

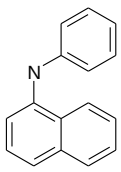
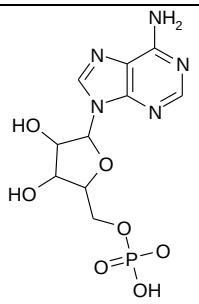
Negative control	*Structure*	*MW (Da)*	*-2 (% at pH 6.8)*	*-1 (% at pH 6.8)*	*0 (% at pH 6.8)*	*+1 (% at pH 6.8)*
N-phenyl-naphthylamine		219	0	0	0	0
imidazole		68	0	0	40.5	**59.5**
adenosine monophosphate		347	**51.49**	48.45	0.06	0

Table 3-5 Negative controls.

Negative controls were chosen as examples of molecules that charge differently at pH 6.8 and are not known to bind to FKBP12. Charge micro species were calculated using the Marvin pK_a calculator (Chemaxon, http://www.chemaxon.com/demosite/marvin/index.html, June 2008). The dominant species is shown in bold typeface.

3.7.7 Ranking the test set by affinity

To conclude, data from the test set suggested that a K_d of around 150 μM was the lowest affinity ligand that can be seen flying in complex with FKBP12 at a protein ligand ratio of 1:10 on the instruments tested. Single concentration measurements were sufficient to predict a rank order under very mild ESI-MS conditions. Absolute affinities were always underestimated for the test set. Moreover, the factor by which the affinity was underestimated was not consistent. Our results suggested that screening at a high ligand concentration could lead to unreliable rank orders of affinity (above 250 μM). AMP was seen to fly in complex at very high ligand to protein ratios. The major micro species for AMP was predicted to carry a charge of -2 at pH 6.8. Theses results suggest that non-specific binding was not a problem for a charged ligand at low ligand concentrations. Results for compounds tested in other ligand series (section 7.8.3) added weight to hypothesis that a single ligand concentration ESI-MS experiment was not a particularly good fine discriminator of ligands in the low micromolar range. Nevertheless, ESI-MS successfully

discriminated between nanomolar, low micromolar (1 – 100 μM) and higher micromolar affinity ligands (> 100 μM).

3.8 Discussion and conclusions

Titration experiments gave an estimated K_d close to the value measured in solution for ligands with low nanomolar affinity. The lowest affinity ligand seen to fly in complex with FKBP12, SD3, had an affinity of 157 μM. Cone voltage dissociation experiments established 15 V to be the minimum cone voltage required to screen a series of compounds on the ZQ instrument using standards screening ratios. Although this voltage led to dissociation of a fraction of the FLBP12:SD1 complex, a lower voltage made it difficult to rank a series of compounds. The variance in repeat measurements (Table 3-4), and a factor in the significance of the rank of a compound, is a function of cone voltage (gross analyte signal to noise), the ease of ionisation of the ligand (degree of suppression of the gross protein signal) and the affinity of the ligand (signal to noise ratio for the protein complex).

An additional consideration of the ranking of compounds by affinity is the nature of the non-covalent interactions between protein and ligand. This is discussed in depth in the next chapter.

Mass spectrometry has the advantage over many biophysical techniques of directly determining the accurate mass of both protein and ligand. Accurate mass determination of recombinant proteins is particularly useful. Single point experiments in electrospray are relatively information rich. Screening experiments carried out at a single ligand concentration give data on the mass of a ligand, its relative binding affinity/stoichiometry and additionally screen for protein stability. A ligand that induces protein destabilization and subsequent unfolding gives a much broader charge-state distribution with an increased population of higher charged species indicative of unfolded protein. ESI-MS has the potential to quickly rank order ligands by a methodology that has great transferability between different protein systems.

4 Chapter 4 Hydrophobic Effects versus Electrostatic Interactions in ESI-MS?

4.1 Introduction

A long standing criticism of investigating protein-ligand interactions by ESI-MS is that the water free environment of a desolvated macromolecular ion is so different to conditions in solution that affinity measurements are compromised. It has been reported that some protein-ligand complexes survive ESI-MS better than others and that the robustness of a complex under a given set of experimental conditions is related to the type as well as the magnitude of the non-covalent interactions between the protein and the ligand (Hernandez and Robinson, 2007). If the affinity of a ligand is to be quantified by the relative proportions of ions representing free protein and protein-ligand complex these observations are of considerable importance. In this section a series of experiments tested the hypothesis: *complexes with a significant enthalpic contribution to the free energy of binding arising from electrostatic interactions dissociate less in electrospray*.

Non-covalent interactions arise from a variety of physical phenomena and exert an attractive force over different ranges and preferred geometries, Table 1-3 (Daniel *et al.*, 2002). It is interesting to note that interaction energies based on charges, dipoles and polarisability are all inversely proportional to the dielectric constant of the medium. Water has a dielectric constant of 78.4 at 25 ^{o}C; estimates for the interior of a protein vary between 1 and 20 (Karp *et al.*, 2007). It would therefore seem reasonable to suppose that electrostatic interactions, between protein and ligand, are strengthened when solvent is removed in the desolvation step of electrospray. Conversely, one would expect hydrophobic effects to be less important under the same conditions. When considering protein complexes in solution, the hydrophobic effect is central to explaining the favourable free energy change on the removal of a hydrophobic ligand from water into a hydrophobic pocket. An additional important consideration is that after desolvation the complex is not in equilibrium with the free protein and ligand; dissociation may occur but association is very unlikely. Such arguments have been used to question the suitability of ESI-MS as a technique for screening small molecules against a target protein (Yin *et al.*, 2008). In a study by Robinson *et al.* (1996) a series of non-covalent complexes between acyl CoA binding protein and acyl CoA derivatives differing by the length of the hydrophobic acyl chain were investigated by ESI-MS. The

fraction of protein seen in complex with ligand in ESI-MS showed poor correlation with ligand affinity in solution. This was rationalized as the increased affinity of the longer chain ligands in solution, due to hydrophobic effect, did not contribute to improving affinity in the gas–phase. It has therefore been proposed that charged ligands, capable of making strong electrostatic interactions, will appear too highly ranked or as false positives in an ESI-MS screen, while complexes where hydrophobic effects are important could be under valued. Complexes of low affinity ligands less reliant on electrostatic interactions may not survive electrospray at all. One could argue that these effects rather than a disadvantage could, when combined with other biophysical and biochemical data, allow one to dissect the relative contributions of different components of the free energy of binding and therefore aid the drug design process.

A metric is required to follow the dissociation of a complex in electrospray. A study by Rogniaux *et al*., (1999) investigating series of aldose reductase complexes found that VC50 (cone voltage required to give 50 % dissociation of the complex) correlated with electrostatic and hydrogen bond interactions (calculated from x-ray structures) and not IC_{50}. In this study the cone voltage induced dissociation profiles for a range of ligands of FKBP12 were examined for a correlation of the VC50 with the affinity and thermodynamic profile of the ligand measured in the solution by ITC. ITC was used to measure affinity in solution and to dissect the free energy of binding into enthalpic and entropic terms (thermodynamic profile). Increasing the cone voltage in ESI-MS raises the speed of macromolecular ions entering the low pressure sector of the instrument. This increases the frequency and kinetic energy of molecular collisions and can result in dissociation of the complex. The aim of this series of experiments was twofold. The primary aim was to establish the sensitivity of ESI-MS over a full range of cone voltage and to define a default cone voltage for screening experiments. The more ambitious aim was to test the hypothesis; *complexes with a significant enthalpic contribution to the free energy of binding arising from electrostatic interactions dissociate less in electrospray*. Could this information complement structure activity relationships derived from measurements in solution and docking studies to predict ligand binding modes if a crystal structure were not available?

4.2 Cone voltage induced dissociation

4.2.1 Materials and Methods

Analytes were dissolved in 20 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). Protein-ligand ratios were chosen so that the majority of the protein was in complex in solution. The ligands investigated by cone voltage induced dissociation are described in Table 4-1. Ratios were calculated using from the law of mass action and K_d values from ITC/literature measured in solution.

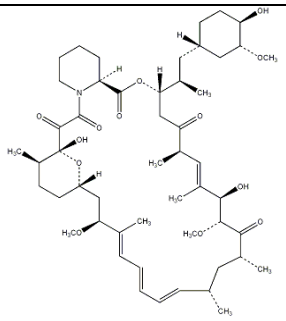
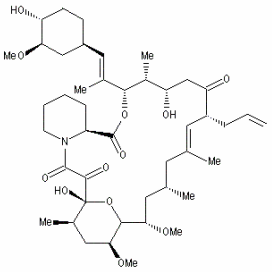
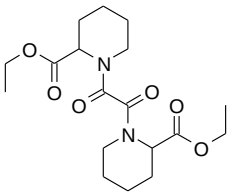
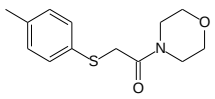
Ligand	*MW (Da)*	*K_d (μM) in solution*	*Structure*	*Reference*
rapamycin	914.1	0.0002, 0.0078		(Bierer *et al.*, 1990; Wear and Walkinshaw, 2007; Wear *et al.*, 2007b)
FK506	804.0	0.0004, 0.001		(Bierer *et al.*, 1990)
KB_61D	368.4	6.6‡/2.1†		Top hit in a series of compounds synthesised to target immunophilins; the KB series. Described in section 7.7.2
SD1	251.1	6.8‡		(Stebbins *et al.*, 2007)

Table 4-1 Ligands investigated by cone voltage induced dissociation

K_d is quoted from the literature for rapamycin and FK506; values for KB2_61D and SD1 were determined by ‡ITC (25 ^{o}C) and †PPIase assay (4 ^{o}C) as part of this study.

The analyte was equilibrated on ice for 5 minutes prior to injection and directly infused into the instrument (Micromass ZQ, Waters) at a flow rate of 10 $\mu l \cdot min^{-1}$ using the instrument syringe pump. Data acquired was an average of 59 scans between 1000 and 2000 m/z, background subtracted and smoothed in Micromass MassLynx V4.1 (Waters). The concentration of the protein-ligand complex was calculated from the ratio of intensity of complexed to uncomplexed protein in the 8+ charge-state pair using the method described in Section 3.6.3. Protein-ligand complex was dissociated by increasing the cone voltage on the instrument between 10 and 120 V. Above this voltage protein fragmentation was observed. The maximum voltage in a given experiment was ligand dependent. The data was recorded for at least 20 V above the voltage where complex could no longer be observed.

The cone voltage required to give 50 % dissociation of the complex (VC50) was estimated from the least squares fit to a generalized reverse sigmoid curve (equation 13) using KaleidaGraph version 4.03 (Synergy Software).

$$PL = \frac{PL_{max} \cdot VC50^{n}}{VC50^{n} + V^{n}} \qquad \textbf{Equation 13}$$

Where: PL is the protein-ligand complex concentration calculated from equation 9, PL_{max} the maximum value of the curve, VC50 the cone voltage required to give 50 % dissociation of the complex, V the cone voltage and n a constant (the steepness of the curve).

4.2.2 Cone voltage induced dissociation profiles

Figure 4-1 shows the cone voltage dissociation profiles for 4 ligands of FKBP12 (affinity range: 0.2 nM to 6.6 μM).

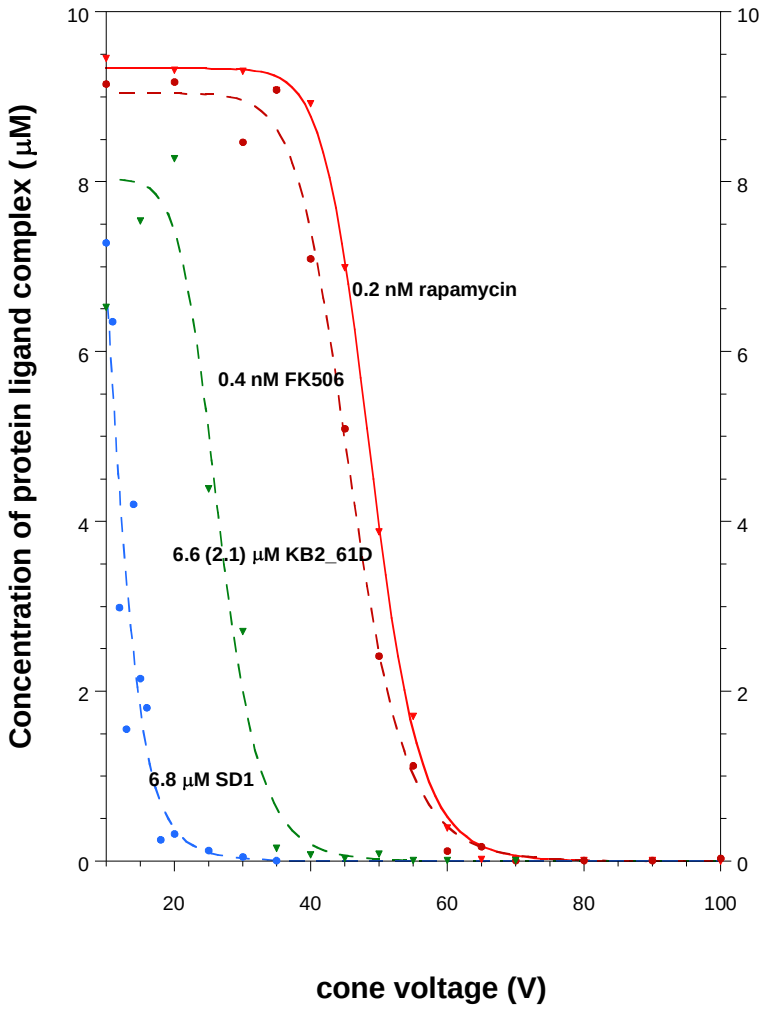


Figure 4-1 Cone voltage induced dissociation.

Cone voltage induced dissociation profiles for complex of FKBP12 and a range of ligands. 9.5 μM FKBP12; 10 μM rapa/FK506 or 100 μM KB2_61D/SD1; dissolved in 20 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). Micromass ZQ instrument, and default parameters, cone voltage 9 to 100 V. Data was acquired over 59 scans; background corrected and smoothed in MassLynx V4.1 (Waters). $[PL]_{\mu M}$ was calculated from the 8+ charge-state pair. K_d is quoted from the literature for rapamycin and FK506 (Bierer *et al.*, 1990; Stebbins *et al.*, 2007; Wear *et al.*, 2007b), the value for KB2_61D and SD1 were determined as part of this study by ITC at 25 °C. The affinity value for KB2_61D in brackets is the value obtained by PPIase assay at 4 °C.

Ligand	*MW (Da)*	*K_d (μM) in solution*	*VC50 (V)*	*Reference*
rapamycin	914.1	0.0002, 0.0078	48.8	(Bierer *et al.*, 1990; Wear *et al.*, 2007b)
FK506	804.0	0.0004, 0.001	45.8	(Bierer *et al.*, 1990; Connelly and Thomson, 1992)
KB2_61D	368.4	6.6‡/2.1†	27.1	EAB Section 7.8.2
SD1	251.1	6.8‡	11.7	(Stebbins *et al.*, 2007), EAB Section 6.4.2

Table 4-2 Correlation of VC50 and solution K_d.

The VC50 was estimated from the least squares fit to a generalized reverse sigmoid curve (equation 13) to the data presented in Figure 4-1, KaleidaGraph version 4.03 (Synergy Software). K_d is quoted from the literature for rapamycin and FK506, values for KB2_61D and SD1 were determined by ‡ITC (25 ^oC) and †PPIase assay (4 ^oC) as part of this study.

Table 4-2 shows that for the compounds tested the ligands of nanomolar affinity for FKBP12 had a much higher VC50 than the micromolar affinity ligands. FK506 and rapamycin are large ligands that bind to FKBP12 with multiple hydrogen bonds and hydrophobic interactions. Interestingly, compounds KB2_61D and SD1 have similar K_d in solution (ITC measurements in this study) however, SD1 was by far the more fragile in electrospray. Dissociation profiles for rapamycin and FK506 have a "horizontal" section, between 10 and ~30 V, where there was less influence of cone voltage on the fraction of complex retained (Figure 4-1). The presence of this section was less clear for KB2_61D, interpretation being more difficult as signal to noise ratios were poor at very low cone voltages. This data suggests that 15 V is close to the minimum value for screening experiments on the Micromass ZQ instrument. Results were in agreement with data obtained on the Micromass ZMD instrument (Figure 3-6), VC50 = 45.1 V for FK506.

4.3 VC50 and H-bonding from crystal structure

It is important to consider any thermodynamic analysis in a structural context. The difference in number of hydrogen bonds between FK506 and rapamycin and FKBP12 was investigated as a simple metric to compare the electrostatic component of non-covalent interactions. Table 4-3 lists hydrogen bond distances (2.5 to 3.3 Å) seen in crystal structures of FKBP12 in complex with FK506 and rapamycin (Figure 4-2 (A)). A hydroxyl group on the cyclohexyl moiety of rapamycin makes an additional hydrogen bond to FKBP12, when compared to that of FK506, due to the closer proximity of Gln53. An early structure of the FKBP12-FK506 complex from Schreiber's group predicted a water mediated hydrogen bond between the equivalent hydroxyl on FK506 and Gln53 (Van Duyne *et al.*, 1991). Not all deposited structures report this water. Rapamycin is a more massive ligand than FK506 and there are additional van der Waals contacts with the 50's loop region of FKBP12 in the region of Glu54. It is important to note that the portions of FK506 and rapamycin that occupy the bulk of the active site are almost identical and are closely aligned in the superimposed structures (Figure 4-2 (B)). The increased VC50 for the rapamycin complex is consistent with the hypothesis that complexes with a larger number of electrostatic interactions require more energy to dissociate in the gas phase. The crystal structure of the FKBP12-FK506 complex is more ordered than that of the FKBP12-rapamycin complex. The entropic penalty for increased structural order has been proposed as contributing to the reduced affinity of FK506 for FKBP12 compared to rapamycin.

Atom: ligand	*Atom : protein*	*Type*	*Distance (Å)*	*Distance (Å)*	*Atom : ligand*	*Atom : protein*	*Type*
FK506:O6	A:37:ASP:OD2	H-bond	2.78	2.61	**rapa:O6**	A:37:ASP:OD1	H-bond
				2.64	**rapa:O13**	**A:53:GLN:O**	**H-bond**
FK506:O10	A:54:GLU:O	H-bond	2.69	2.67	**rapa:O10**	A:54:GLU:O	H-bond
FK506:O2	A:56:ILE:N	H-bond	2.80	2.93	**rapa:O2**	A:56:ILE:N	H-bond
FK506:O3	A:82:TYR:OH	H-bond	2.76	2.66	**rapa:O3**	A:82:TYR:OH	H-bond

Table 4-3 A comparison of probable hydrogen bonds and FK506 and rapamycin to FKBP12 derived from x-ray structure.

FKBP12-FK506: 1FKJJ.pdb, FKBP12-rapamycin: 2DG3.pdb (Fulton *et al.*, 2003). Inter-atomic distances (heavy atom to heavy atom distances in angstrom) were calculated using the program PyMol V1.02 (DeLano Scientific). The additional hydrogen bond in the interaction with rapamycin is shown in bold.

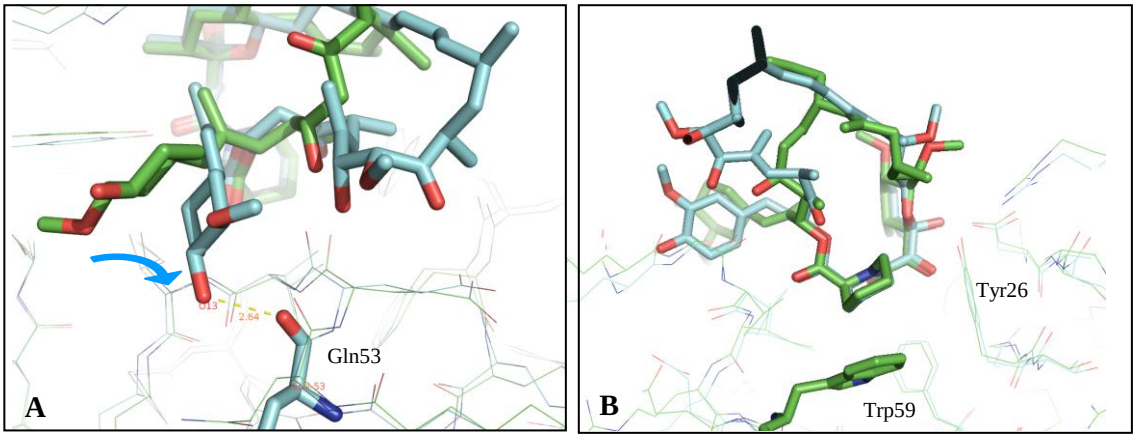


Figure 4-2 Details of x-ray structures of FKBP12 in complex with rapamycin and FK506.

(A) The blue block arrow shows the additional hydrogen bond is gained on rapamycin binding to FKBP12 (2DG3.pdb) when compared to the structure of FK506 in-complex with FKBP12 (1FKJ.pdb). This is the result of the hydroxyl group on the cyclohexyl ring being closer to Gln53 in the rapamycin structure. The putative hydrogen bond is shown as a dashed line, distance is labelled in angstrom. (B) The common structural elements of FK506 and rapamycin occupy very similar positions in the interior of the hydrophobic pocket. Trp59 is shown highlighted; this forms the "floor" of the pocket. Structures were superimposed in PyMol V1.02 (DeLano Scientific).

4.4 Correlation of VC50 and binding mode from docking experiments

4.4.1 Docking SD1 into FKBP12

It was not possible to crystallise FKBP12 in complex with SD1 or KB2_61D in this study. Docking experiments were carried out for compounds SD1 and KB2_61D using FlexX (Rarey *et al.*, 1996) and LIDAEUS (Wu *et al.*, 2003) to generate binding mode hypotheses. Docking strategies employed in this project are described in Chapter 7.

Modelling of SD1 into the active site of FKBP12 was carried out using the docking programs LIDAEUS and FlexX, and a protein template from a high resolution x-ray of the FKBP12-rapamycin complex (2DG3.pdb, 1.7 Å resolution). The two clusters of high scoring poses suggested by FlexX are illustrated in Figure 4-3. Poses were analysed for correlation with structure activity relationships derived from NMR data. The SAR analysis is described in 4.4.2. The top scoring FlexX pose for SD1 fitted well with the SAR analysis. This pose predicted one hydrogen bond between the amide oxygen of SD1 and the hydroxyl of Tyr82 (mode 1, Figure 4-4). The top scoring example in an alternative cluster of poses is also shown in Figure 4-4 (mode 2). In mode 2 the ligand lies less deeply in the active site, the morpholine ring makes fewer van der Waals contacts with Trp59. Docking using LIDAEUS and a resolution of 0.06 Å suggested a very similar array of preferred orientations in the active site.

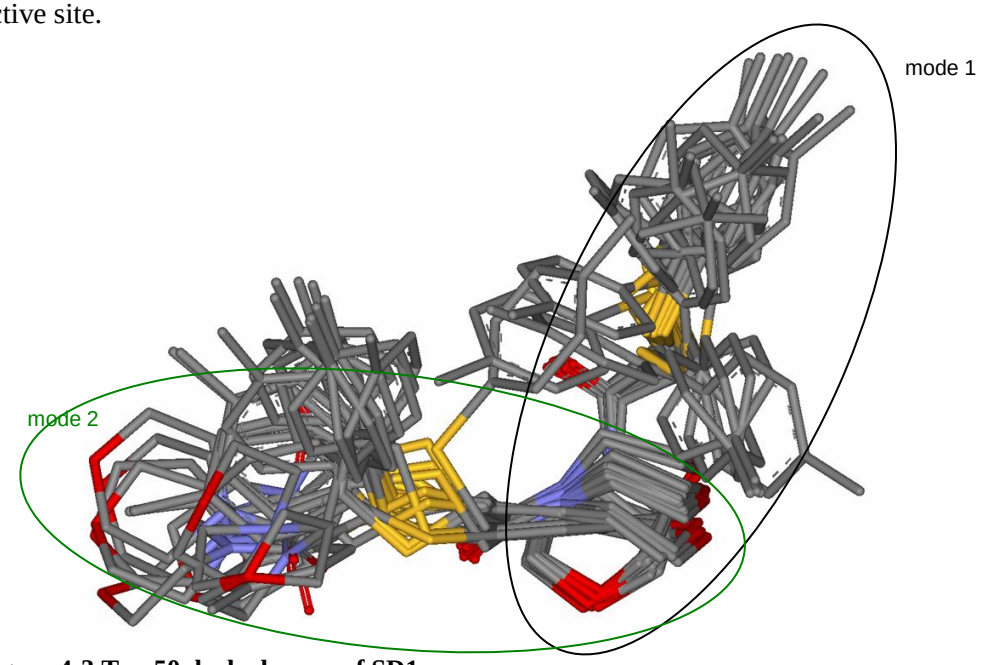


Figure 4-3 Top 50 docked poses of SD1

The top 50 FlexX poses for SD1 were clustered into two regions of the active site. FlexX V3.0.2 (Rarey *et al.*, 1996).

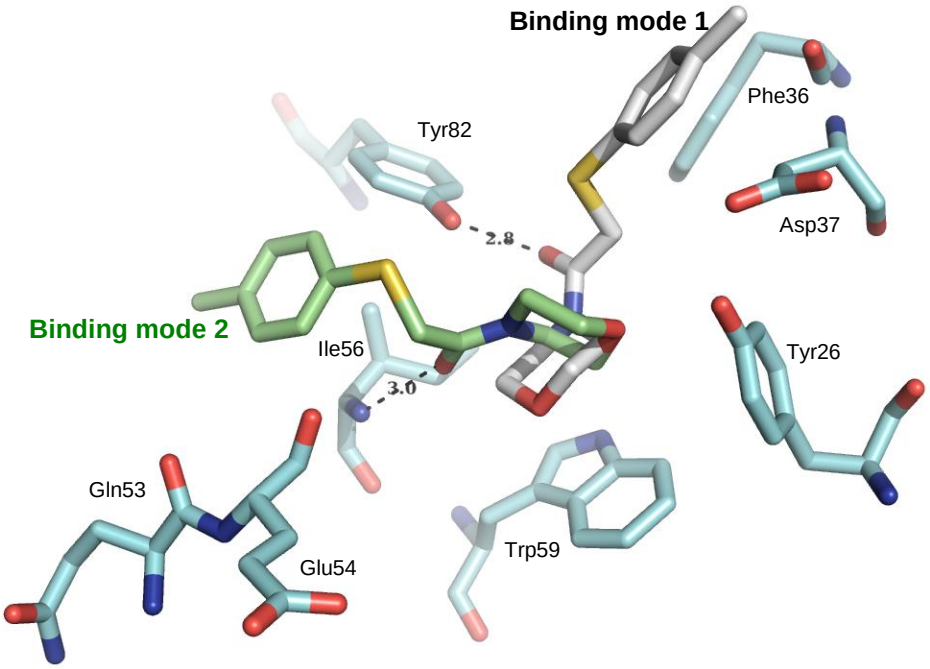


Figure 4-4 Predicted binding modes for SD1 in complex with FKBP12

Binding mode 1 (grey carbon skeleton). The top scoring pose makes a hydrogen bond with the hydroxyl group of Tyr82, the morpholine ring is placed deepest in pocket and the phenyl ring in a hydrophobic trench close to Phe36. This pose fits best with the SAR analysis, 4.4.2.
Binding mode 2 (green carbon skeleton), the ligand makes a hydrogen bond with the proton on the main chain nitrogen of Ile56. Important residues in the active site of FKBP12 are shown with blue carbon skeletons. FlexX V3.0.2 (Rarey *et al.*, 1996).

In Figure 4-5 (A) the top scoring docking solution for SD1 (mode1) has been aligned with the x-ray structure of rapamycin in complex with FKBP12 (2DG3.pdb). Parts (B) and (C) show 2D schematic diagrams of non-covalent interactions for the FKBP12:rapamycin complex and the best scoring pose of SD1.

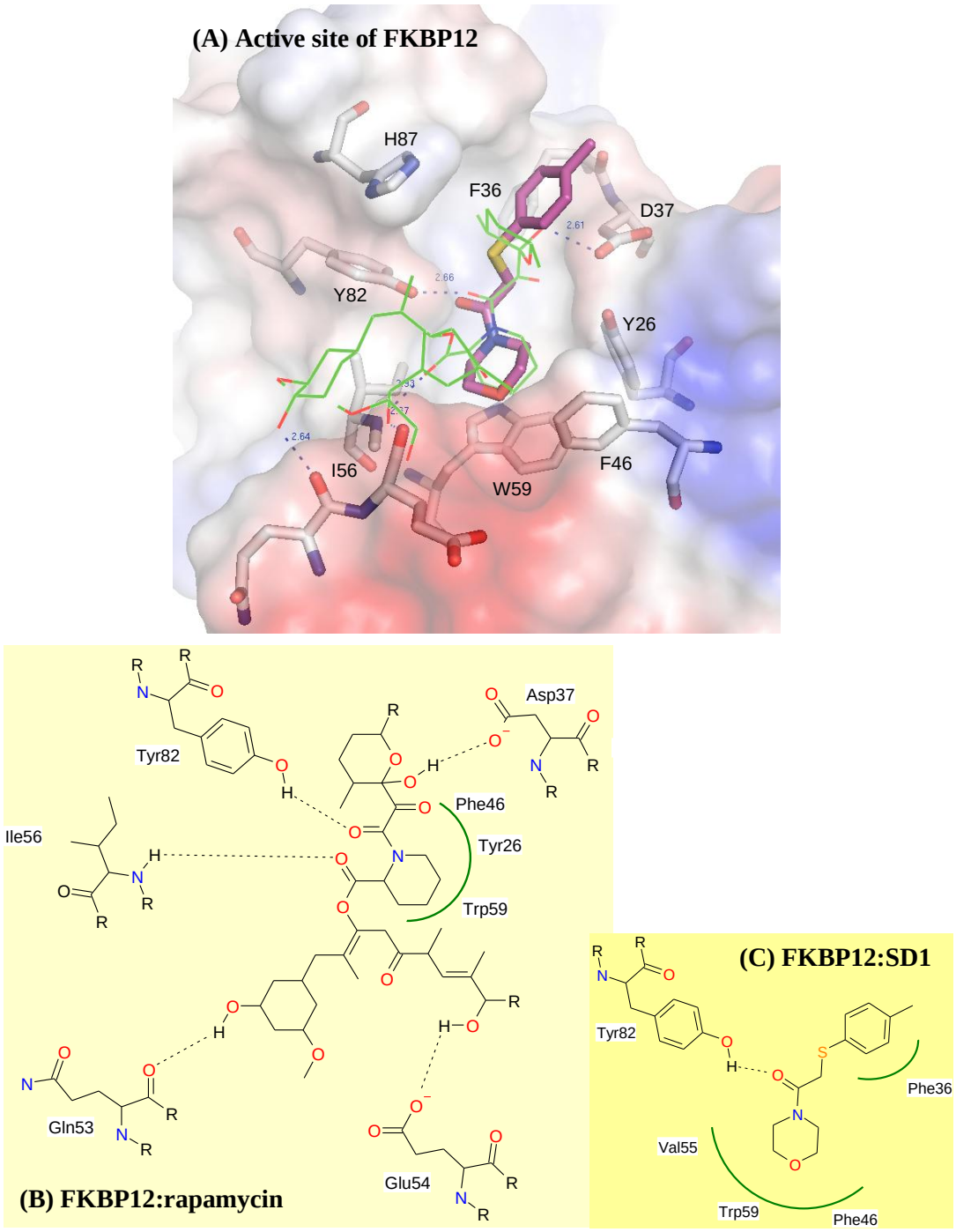


Figure 4-5 Modelling SD1 into the active site of FKBP12.

(A) The highest scoring docked pose of SD1 is shown with a purple carbon skeleton. Rapamycin is shown as green carbon sticks; hydrogen bonds from rapamycin to FKBP12 are shown as blue dashed lines and labelled in angstrom. The solvent exposed loop of rapamycin has been hidden for clarity. Key protein residues are highlighted. An electrostatic surface was generated to illustrate the geometry of the active site; PyMol V1.02 (DeLano Scientific). (B) 2D Schematic representation of the non-covalent interactions between rapamycin and FKBP12 (2DG3.pdb). (C) 2D Schematic representation of the top scoring docked pose of SD1. Hydrogen bonds are shown as dashed lines; selected hydrophobic interactions are shown as green arcs. Docking was performed in FlexX V3.0.2 using the x-ray structure 2DG3.pdb as a template protein structure (Rarey *et al.*, 1996).

4.4.2 **Structure activity relationships within the San Diego data set**

The small molecules in published crystal structures of FKBP12 complexes are not structurally diverse (Table 1-4). The majority share a core motif with rapamycin and FK506 and bind in a very similar mode (Figure 1-9). Crystal structures were only available for two members of the test set; it was felt that structure activity relationship (SAR) analysis for the San Diego data set would be helpful in devising binding mode hypotheses for the compounds lacking an x-ray structure. The San Diego data set comes from a high-throughput NMR screen against FKBP12; it is described in section 3.7.1. This data set was particularly valuable as it contained experimental data on actives and in-actives. The library was complied with knowledge of the target and contained clusters of relatively close analogues. Of the 3775 compounds 44 were reported to be active ($K_d \leq 500$ μM). The data set was analysed for chemical similarity. A 2D chemical fingerprint was written for each compound and an inter-compound pair-wise comparison made by means of a Tanimoto similarity coefficient (Hert *et al.*, 2006; Eckert and Bajorath, 2007). Many of the higher affinity compounds in the San Diego data set were structurally similar to compound SD1. These compounds consist of a morpholine ring with an amide bond and a $–CH_2$-S- linker to an aromatic ring. Figure 4-6 summarises the finding of the SAR study, using SD1 as the structure for comparison. It is worth noting that uncertainties were not listed in the NMR data.

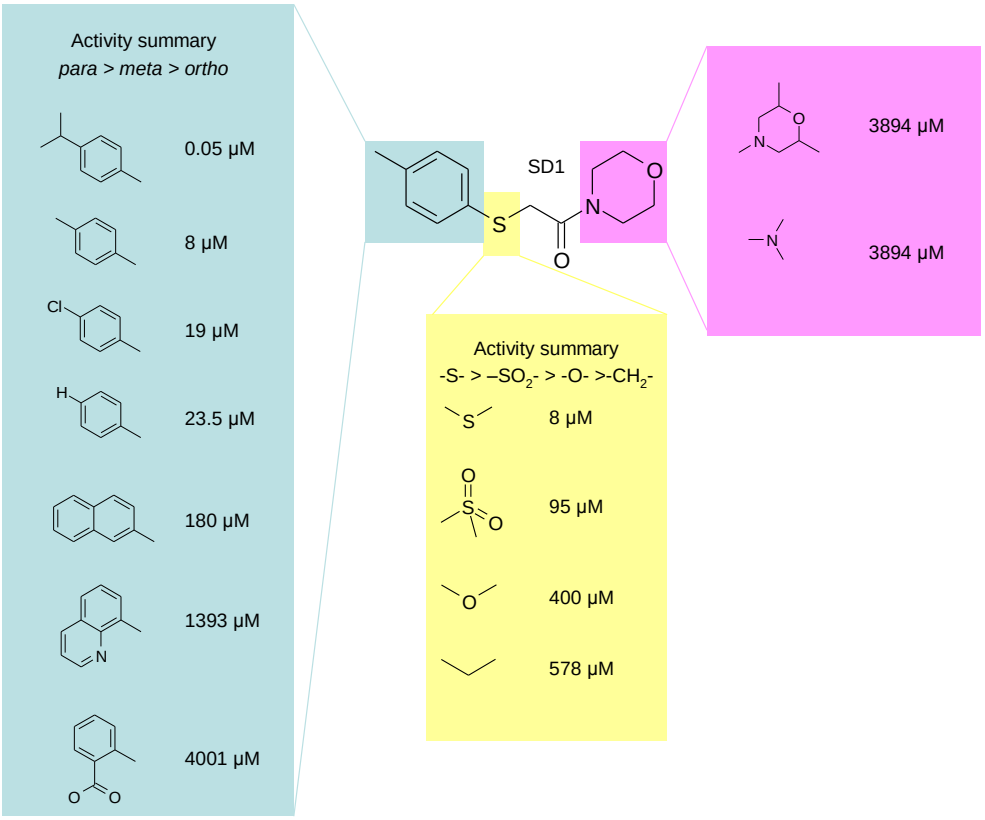


Figure 4-6 Structure activity relationships for the SD1 cluster.

The blue box illustrates the effect of substituting the phenyl group of SD1. Replacing sulfur in the linker is shown in the yellow box and replacement of the morpholine ring in the pink box. SDCCG-A014-FKBP12-NMR, http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=608, September 2008.

4.4.3 A binding mode hypothesis for SD1

Evidence from docking and SAR suggests that the morpholine ring of SD1 mimics the pipecolate ring of rapamycin by making contacts with Trp59 at the base of the hydrophobic pocket (Figure 4-5 (A)). This is supported by the evidence that the addition of two methyl groups *ortho* to the oxygen on the morpholine ring greatly reduced activity. Docking studies indicated that these substitutions make the ring too bulky to occupy the base of the pocket. The amide oxygen is important for activity; shortening the linker between the rings (deletion of the carbonyl group) removed activity. Docking suggested that the amide oxygen may make a hydrogen bond with the hydroxyl of Tyr82, mimicking the hydrogen bond seen with the amide oxygen of rapamycin. Substitution to the aromatic ring in the para position has a negligible effect on activity with the exception of the most active member of the cluster where the substitution of an isopropyl group for the methyl of SD1 in the para position

significantly increased activity. Modelling suggested that allowing the 80's loop region of FKBP12 to move out would accommodate the isopropyl group in the hydrophobic pocket formed by the 80's loop, rotating the ligand slightly from the mode proposed for SD1 (anti-clockwise, consider an axis orthogonal to the plane of the paper through the carbon of the amide, Figure 4-5 (A)). It is worth noting that NMR is not usually considered the best technique for estimating the affinity of ligand in the nanomolar range. There is likely to be a considerable degree of uncertainty in the reported K_d for this particular ligand. Additions to the ring at *ortho* and *meta* positions reduced affinity. This is consistent with the aromatic ring of SD1 lying in a narrow hydrophobic trench adjacent to Asp37 and Phe36. Making the aromatic ring wider by substitution in the *ortho* and *meta* positions would make the ring too broad to fit into this trench. Substitution in the *para* position does not preclude binding in this mode as *para* groups would become solvent exposed. An alternative binding mode for SD1, suggested by docking experiments, proposed that the aromatic ring lies in the hydrophobic trench bound by Gln53 (mode 2, Figure 4-4). This mode suggested that the amide oxygen made a hydrogen bond with the hydrogen of main chain nitrogen of Ile56, mimicking the hydrogen bond between Ile56 and the lactone oxygen of rapamycin. This is less likely as the Gln53 hydrophobic trench is more accommodating than the trench bounded by Asp37 and additions to the ring in the *ortho* and *meta* positions would be less likely to "push" the ligand out of the pocket thus reducing the number of non-covalent interactions. Substituting –S- in the linker region for $-SO_2-$, –O- or $–CH_2-$ reduced activity. These substitutions have the effect of altering the geometry of the molecule and reducing van der Waals contacts.

These observations are dependent on there being no gross changes to protein structure on ligand binding. Evidence from numerous crystal structures of complexes indicate that FKBP12 does not change significantly on ligand binding, but there are localized changes in the loop regions (Van Duyne *et al.*, 1991; Wilson *et al.*, 1995). Changes in this loop region are observed in x-ray structures and are considered to be important most significant for the FKBP12-ligand interface required for the immunosuppressant action of rapamycin and FK506.

4.4.4 **Docking compound KB2_61D**

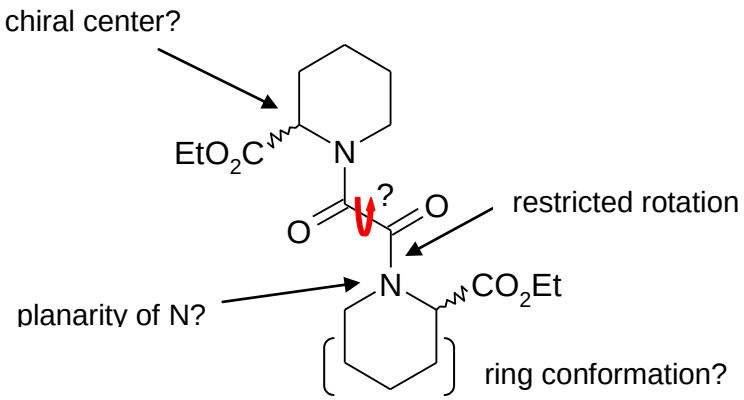


Figure 4-7 Stereochemistry of compound KB2_61D.

Structural elements of compound KB2_61D that make it difficult to predict the isomer/conformer for use in docking experiments.

Proposing a binding hypothesis for KB2_61D is less straightforward than for SD1. The compound was not made by a stereospecific route (use of racemic ethyl pipecolate as a reagent). KB2_61D contains two chiral centers, two amide bonds with restricted rotation (two dominant rotamers) and a di-keto linker where the torsion angle between the two carbonyl groups is difficult to predict (Figure 4-7). The NMR spectrum of the sample was characteristic of a mixture and it proved impossible to predict the relative proportions of each isomer (Dr. Kevin Bailey, University of Manchester). A cursory inspection of the molecule would indicate that KB2_61D has more potential for making electrostatic interactions than SD1. It was decided that docking every diastereomer and a wide range of likely conformers would probably not be helpful. A non-exhaustive search exploring various degrees of freedom did not produce a clear cut result. An alternate strategy was adopted. KB2_61D is a symmetrical analogue of the pipecolate ring and the di-keto moieties found in FK506 and rapamycin. The conformation of rapamycin and FK506 have been discussed in the literature (Wilson *et al.*, 1995). In the bound forms, both FK506 and rapamycin have orthogonal carbonyl groups in the di-keto moiety (O-C-C-O torsion ~ 90^{o}) and the amide in the *trans* form. It was decided to model compound KB2_61D on the pipecolate ring and di-keto group of rapamycin. This was done by fixing the chirality of one pipecoate ring to that of rapamycin (S). The torsion angles were set to similar values to those in rapamycin and then minimized using a MFF94 force field in SYBYL (Tripos Software). Minimisation gave almost planar nitrogen in the pipecolate ring. Both forms of the second stereogenic center

were docked using FlexX V3.0.2 (Rarey *et al.*, 1996). The best solution (lowest energy) is shown in Figure 4-8.

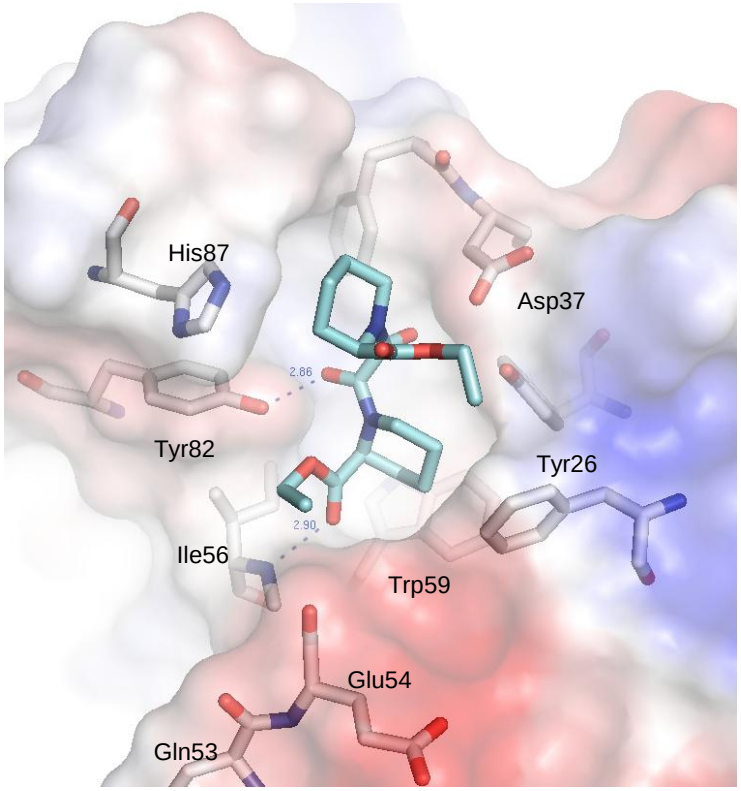


Figure 4-8 Docking pose for compound KB2_61D.

The best pose for compound KB2_61D ((S),(R) form). Two proposed hydrogen bonds are shown as blue dashed lines, labelled in angstrom. FlexX V3.0.2 (Rarey *et al.*, 1996).

This pose fits with SAR derived from the KB2 series (Figure 7-18, SAR showed that an ortho substituted ester increased activity as did the second pipecolate ring when compared to a non-symmetrical analogue, 62D). This pose predicted 2 hydrogen bonds; Ile56 –NH to the carbonyl on one pipecolate ring and Tyr82-OH to an amide carbonyl of the same ring (first carbonyl of the di-keto moiety). The second carbonyl of the di-keto moiety occupied the same small hydrophobic pocket adjacent to Phe99 as the analogous group in rapamycin/FK506.

The docking results suggested that KB2_61D makes more electrostatic interactions than SD1. This would be in agreement with the hypothesis, *complexes with a predominance of electrostatic interactions survive electrospray better*. However, there is a caveat to this conclusion. KB2_61D is a mixture of diastereomers. If the synthesis shows no chiral selectivity, only 25 % of the compound is the diastereoisomer shown in Figure 4-8. However, docking predicts the epimers (S),(R) and (S),(S) both fit reasonably well into the

active site. The chirality of the pipecolate positioned deepest in the active site was most important for the probability of making two hydrogen bonds with the protein. We would suggest that not all diastereomers of KB2_61D are equally active and that having one pipecolate in the (S) form improves activity relative to the (R) form. Thus, the solution K_d is not representative of the most active diastereomers. In addition, it makes the assumption that rotamers are inter-convertible. Inter-conversion of rotamers is particularly interesting in the context of FKBP12 as a peptidyl-prolyl isomerase. To further resolve these issues it would be necessary repeat the affinity experiments with a single diastereomers of KB2_61D also to crystallise each respective complex. A first step might be to try to separate diastereomers on a C18 reverse-phase HPLC column.

4.5 Exploring collision induced dissociation in different protein systems

4.5.1 Introduction

Preliminary data from the collision induced dissociation of FKBP12 complexes supported the hypothesis: *complexes with a significant enthalpic contribution to the free energy of binding arising from electrostatic interactions dissociate less in electrospray*. To further this study required additional experimental data for ligands with nano to low micromolar affinity for FKBP12. As additional ligands were not available, two different model protein systems with low micromolar affinity ligands were investigated. The proteins selected had similar molecular weight and solvent exposed surface area to FKBP12, but structurally and chemically different ligands. It was important to select proteins that flew in electrospray with sufficient charge to be observable in the m/z range of the instrument. The proteins selected were all weakly basic and had similar pIs.

Protein	**Structure**	**Resolution (Å)**	**MW (Da)**	**Surface Area ($Å^2$)**	**pI_{calc}**	**Ligand**	***K_d (μM) (ITC, 25 °C, pH 6.8)***
FKBP12	2DG3.pdb	1.7	13348	6091	8.79	SD1	6.8
HEWL	1HEW.pdb	1.75	14313	6904	9.32	N, N', N''-Triacetylglucosamine	8
RNaseA	1ROB.pdb	1.6	13690	7010	8.64	2'-cytidine monophosphate	1.5

Table 4-4 Summary protein properties for the model systems.

Molecular weight and pI were calculated from protein sequence using the ExPASy ProtParam server (www.expasy.org/tools/protparam.html, June 2008, (Gasteiger *et al.*, 2005). Surface area was calculated using a probe radius of 1.4 Å and a Connolly surface, PyMol V1.02 (DeLano Scientific). K_d were measured in this study by ITC, pH 6.8, 25 ^oC.

Table 4-4 summarises basic structural descriptors and affinity for FKBP12, hen egg white lysozyme (HEWL) and ribonuclease A (RNaseA) and their respective ligands. Figure 4-9 shows x-ray structures of FKBP12, HEWL and RNaseA with their respective ligands and 2D schematic diagrams of protein-ligand non-covalent interactions.

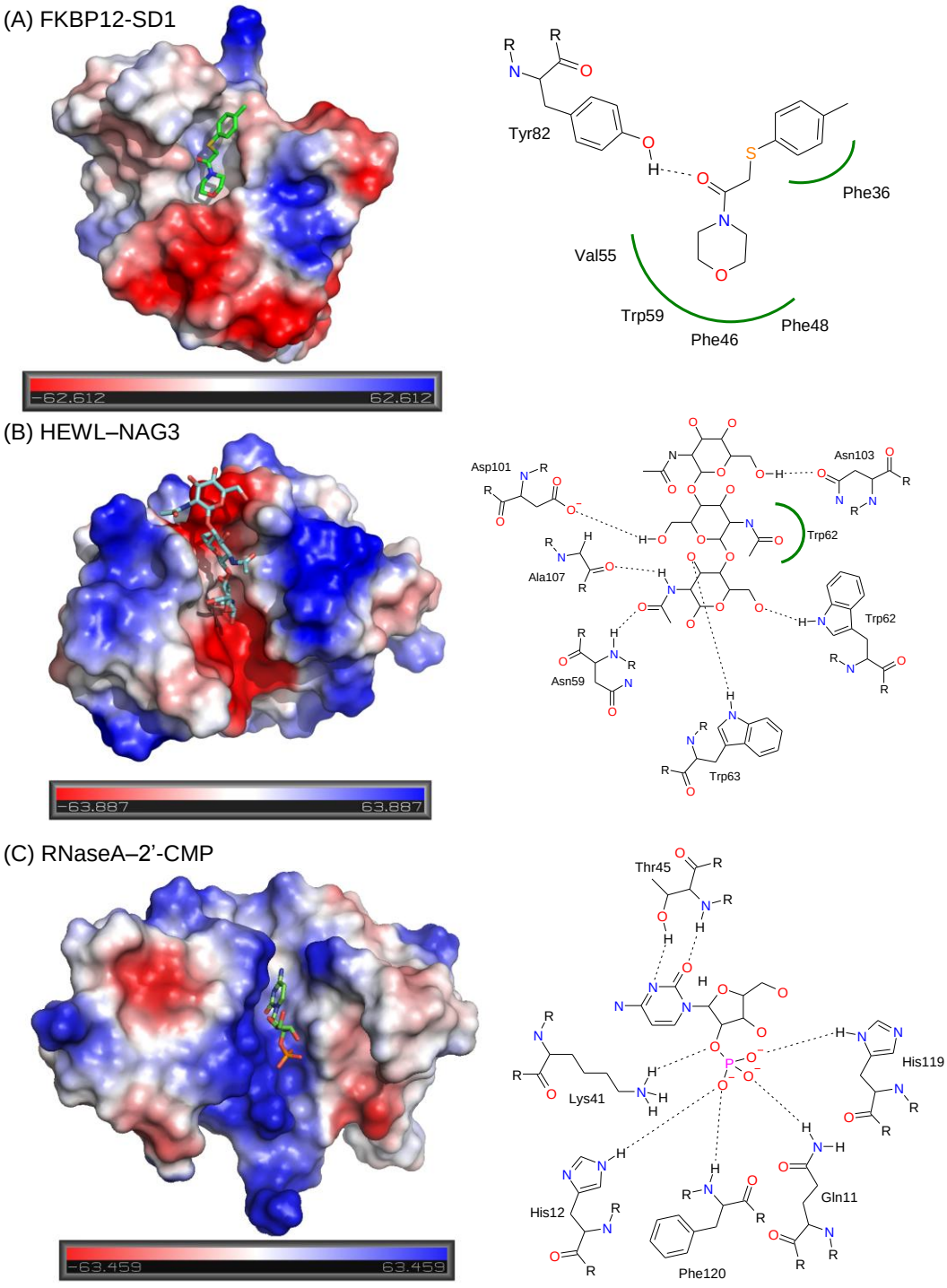


Figure 4-9 Representation of protein-ligand interactions in three model systems.

(A) Top scoring docking pose of SD1 and FKBP12 (2DG3.pdb protein template). Schematic diagram of predicted non-covalent interactions between FKBP12 and SD1. (B) X-ray structure of HEWL and N, N', N''-Triacetylglucosamine (NAG3) and schematic diagram of non-covalent interactions (1HEW.pdb). (C) X-ray structure of ribonuclease A (RNaseA) and schematic diagram of electrostatic interactions (1ROB.pdb). Protein-ligand electrostatic interactions are represented by dashed lines and hydrophobic interactions by solid green arcs. Selected hydrogen atoms are shown for clarity.

The electrostatic surfaces shown in Figure 4-9 illustrate the different nature of the binding sites for the three protein systems. FKBP12 has a predominantly hydrophobic pocket. HEWL has a negatively charged cleft that binds polysaccharide substrates. Aspartic acid and glutamic acid residues are important for the catalytic activity of the enzyme. RNaseA has a positively charged cleft that binds RNA; two histidine residues are important for catalytic action.

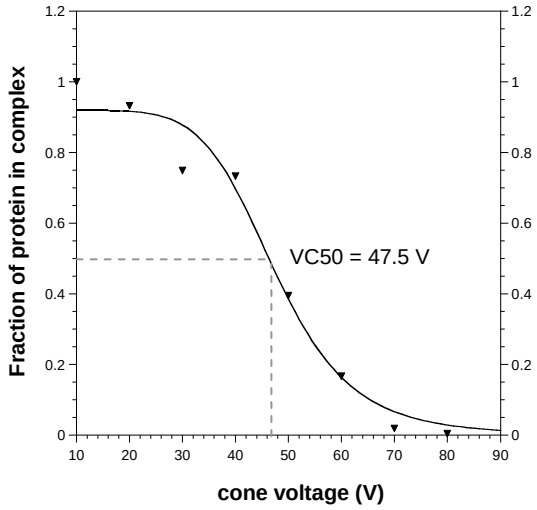
4.5.2 **Hen egg-white lysozyme and N, N', N''-Triacetylglucosamine**

N, N', N''-Triacetylglucosamine (NAG3) is a tri-saccharide that binds specifically to hen egg-white lysozyme (HEWL) in a 1:1 ratio non-covalently bound complex. HEWL functions as a hydrolase, with specificity for β(1-4) glycosidic bonds between *N*-acetylmuramic acid and *N*-acetylglucosamine (NAG). NAG3 does not undergo hydrolysis when bound to the enzyme. High purity HEWL (lyophilised, salt-free) and NAG3 (> 95 %) were purchased from Sigma, UK. HEWL was dissolved in 50 mM ammonium acetate solution, pH 6.8 and NAG3 in water. The HEWL:NAG3 complex has been studied using a variety of electrospray mass spectrometry approaches (Jecklin *et al.*, 2008). NAG3 has been shown to fly in complex with HEWL under conditions that retain active enzyme.

HEWL has theoretical molecular weight of 14313 Da, it flew in electrospray with an apparent mass of 14311.2 Da. The complex of HEWL and NAG3 had an apparent mass of 14938.4 Da, this was an addition of 627.2 Da. 2'-CMP has a mass of 627.6 Da, this suggests that the complex is a 1:1 ratio of protein to ligand and that there were no additional water molecules present for the dominant species. This is interesting as the x-ray structures of HEWL in complex with NAG3 show 2 hydrogen bonded water molecules bridging protein and ligand. These are apparently not required for binding in the gas-phase.

Figure 4-10 (A) shows how the fraction of complex decreased as the cone voltage was raised from 10 to 80 V. The VC50 was 47.5 V; there was no complex present after 70 V. The affinity of NAG3 for HEWL is 8.0 μM at pH 6.8, ITC data, Figure 4-10 (B). Binding was exothermically driven, ΔH = -12.5 kcal·cal^{-1}. However, there was an entropic penalty for the bound state, $T\Delta S$ = -5.5 kcal·cal^{-1}. The ITC data was consistent with literature values; the binding of carbohydrates to proteins is typically enthalpically driven (Garcia-Hernandez *et al.*, 2003). Interestingly, the VC50 is similar to that of the FKBP12-FK506 complex (a much higher affinity complex).

(A) HEWL-NAG3



(B) ITC

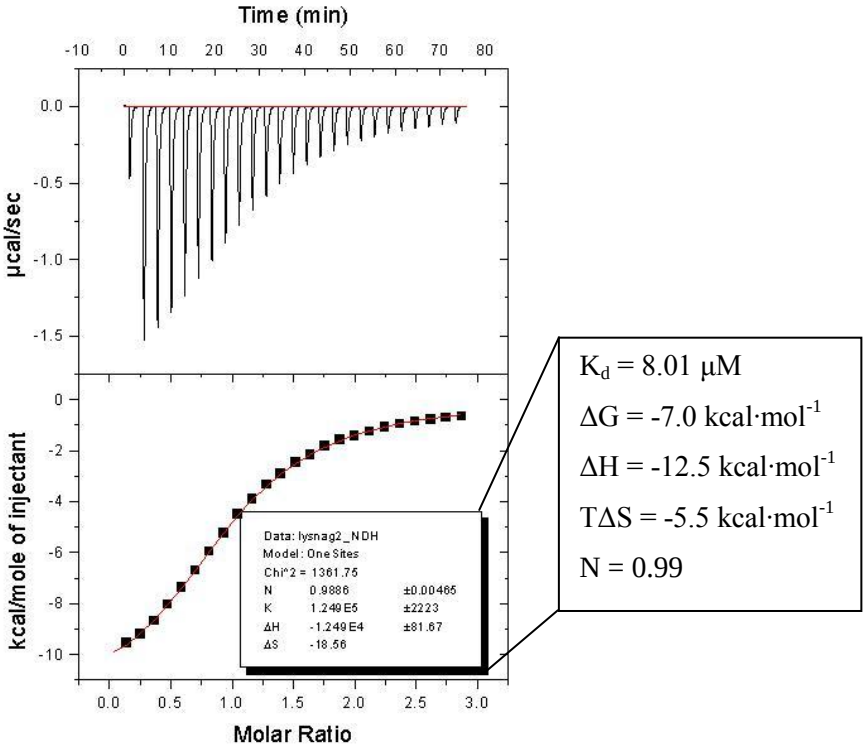


Figure 4-10 Influence of cone voltage on HEWL-NAG3 preserved in ESI-MS

(A) Fraction of complex preserved as a function of cone voltage. The data was normalised to the fraction of complex at a cone voltage of 10 V. 20 µM HEWL, 100 µM NAG3, 20 mM ammonium acetate, pH 6.8. (B) Raw ITC data. 0. 5 mM NAG3 was injected into 1.42 ml, 32 µM HEWL in 25 x 10 µl injections (3 µl 1st injection), 180 s apart at 25 ^{o}C. All reagents were dissolved in 50 mM sodium phosphate, pH 6.8 (Bottom panel) Integrated heat data fit to a standard 1:1 binding model (MicroCal Origin software, V4.0).

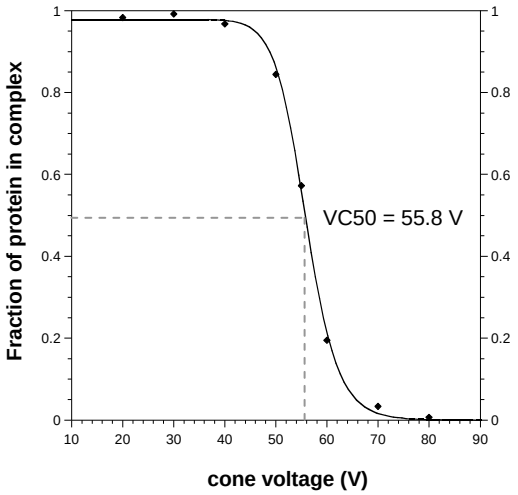
4.5.3 **Ribonuclease A and 2'-Cytidine monophosphate**

2'-Cytidine monophosphate (2'-CMP) is a nucleotide mimetic that binds specifically to Bovine pancreatic ribonuclease A (RNaseA) in a 1:1 ratio non-covalently bound complex (Lisgarten *et al.*, 1993). RNaseA is an endonuclease, it catalyses the cleavage of phosphodiester bonds of ribonucleotides, the mechanism of action is well understood. This enzyme is known to be particularly stable and has been extensively used as a model protein system. Cytidine monophosphate (2'-CMP) acts as an inhibitor of the natural substrate of this enzyme, the phosphate group is not cleaved by the enzyme because of the 2', rather than 3' link between the phosphate group and the ribose ring. The enzyme was purchased as a lyophilised powder and dissolved in 50 mM ammonium acetate buffer, pH 6.8 (Ribonuclease A from bovine pancreas, Type XII-A, ≥90 % (SDS-PAGE), Sigma, UK). 2'-CMP was dissolved in methanol (Sigma, UK).

RNaseA has theoretical molecular weight of 13690.2 Da, it flew in electrospray with an apparent mass of 13686.8 Da. The complex of RNaseA and 2'-CMP had an apparent mass of 14009.1Da, this was an addition of 322.3 Da. 2'-CMP has a mass of 323.2 Da, this means that the complex is a 1:1 ratio of protein to ligand and that there were no additional water molecules present for the dominant species flying in ESI-MS. This is consitent with the x-ray data (1ROB.pdb) where no water molecules are seen to bridge between the ligand and the protein.

Figure 4-11 (A) shows how the fraction of complex decreased as the cone voltage was raised from 10 to 80 V. The VC50 was 55.8 V; there was no complex observed after 80 V. The affinity of RNaseA for 2'-CMP 1.5 μM at pH 6.8. Binding was exothermically driven, ΔH = -16.9 kcal·cal^{-1}. However, there was an entropic penalty for the bound state, $T\Delta S$ = -8.7 kcal·cal^{-1}. ITC data (Figure 4-11) showed ligand binding to be exothermic however there was a very significant entropic penalty on binding. ITC data was consistent with literature values (Spencer *et al.*, 2002).

(A) RNaseA-2'-CMP



(B) ITC

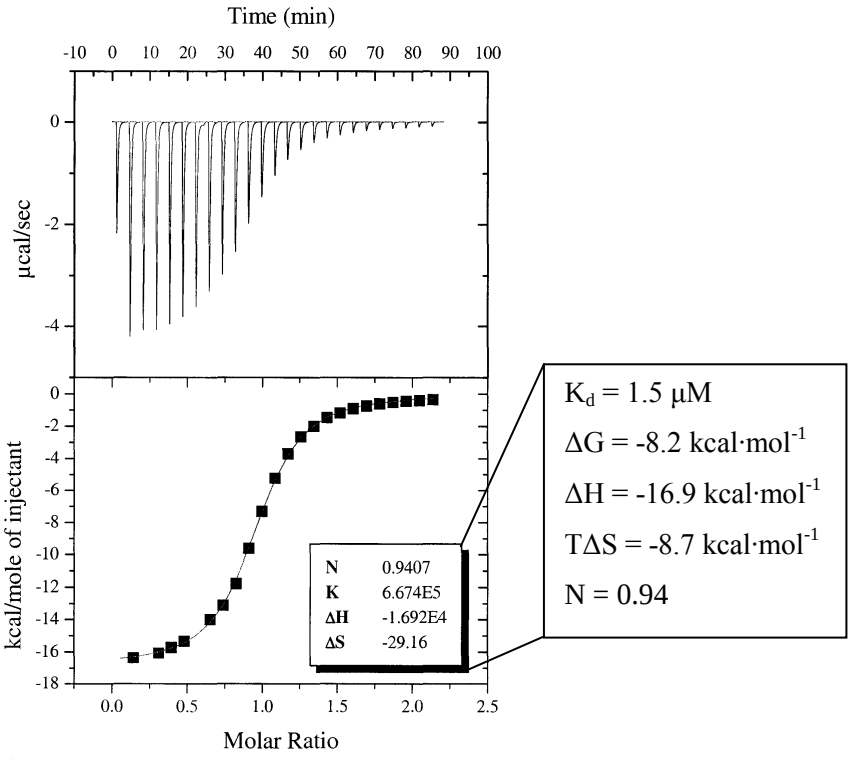


Figure 4-11 Influence of cone voltage on RNaseA-2'-CMP complex preserved in electrospray.

(A) Fraction of complex preserved as a function of cone voltage. The data was normalized to the fraction of complex at a cone voltage of 10 V. 42.5 µM RNaseA, 75 µM 2'-CMP, 20 mM ammonium acetate, pH 6.8. (B) Raw ITC data for 2'CMP injected into RNaseA All reagents were dissolved in 50 mM sodium phosphate, pH 6.8, 25 ^{o}C. (Bottom panel) Integrated heat data fit to a standard 1:1 binding model (MicroCal Origin software, V4.0). (ITC data courtesy of Dr Samantha Rutherford, Organon).

4.5.4 **Correlation between VC50 and thermodynamic data**

Protein	*ligand*	*Hydrogen bonds*	*VC50 (V)*	*ΔH (kcal·mol^{-1})*	*TΔS (kcal·mol^{-1})*	*ΔG (kcal·mol^{-1})*	*Stoichiometry N*	*K_d (μM)*
FKBP12	SD1	*1*	**11.7**	**-6.09**	0.96	-7.05	0.96	6.8
FKBP12	KB2_61D	2	**27.1**	**-11.7**	-4.6	-7.1	0.89	6.6
FKBP12	FK506	4	**45.8**	**(-14.2)**	(-2.3)	(-12.3)	NA	0.0004
FKBP12	rapamycin	5	**48.8**	**(-16.4)**	(-3.49)	(-12.7)	NA	0.0002
HEWL	NAG3	6	**47.5**	**-12.49**	-5.53	-6.96	0.99	8.0
RNaseA	2'-CMP	7	**55.8**	**-16.92**	-8.69	-8.23	0.94	1.5

Table 4-5 Correlation between VC50 and thermodynamic data.

VC50 is the cone voltage required to give 50 % dissociation of the complex. VC50 was determined by least squares fit of cone voltage dissociation data to equation 13 using KaleidaGraph version 4.03 (Synergy Software). Thermodynamic data is from this study except values in brackets which are reproduced from (Connelly and Thomson, 1992). The number of predicted hydrogen bonds is from x-ray data, except SD1 which is from the model described in section 4.4.3 (shown in italics). K_d for F506 and rapamycin are from (Bierer *et al.*, 1990)

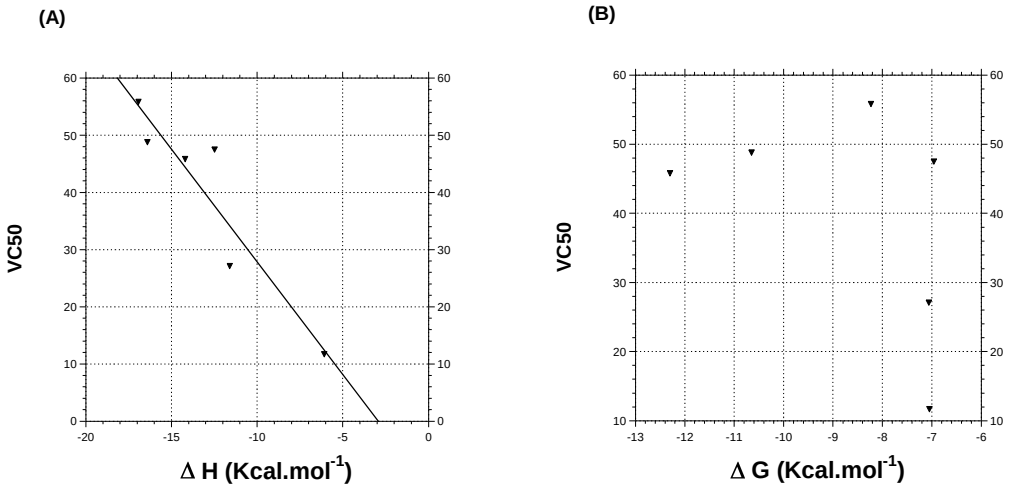


Figure 4-12 Correlation between V50 and enthalpic contribution to binding

There is a correlation between VC50 and the enthalpic contribution to the Gibbs free energy change for the protein ligand interaction.

Table 4-5 shows that there is a correlation between VC50 and the number of hydrogen bonds between the ligand and complex predicted from an x-ray structure/molecular model. A correlation is seen between the VC50 and ΔH measured by ITC, Figure 4-12 (A). There is no correlation between VC50 and ΔG, Figure 4-12 (B). These results, although limited in scope, confirm the hypothesis that complexes with a predominance of electrostatic interactions

survive electrospray better. The RNaseA-2'-CMP and HEWL-NAG3 complexes both form multiple hydrogen bonds between protein and ligand. The binding of both proteins to their respective ligands is enthalpically driven with a significant entropic penalty.

4.6 Discussion and conclusions

4.6.1 Minimum cone voltage

Setting cone voltage in ESI-MS is always a compromise between sensitivity and good signal to noise ratio. Every protein-ligand complex has an optimum cone voltage for characterisation of the interaction by MS; for example, estimating the affinity of the interaction by the titration method. This study confirmed the finding in Chapter 3 (3.7) that 15 V was the lowest cone voltage that could be used in screening experiments on the Micromass ZQ instrument. Cone voltage dissociation profiles proved to be a useful tool to quantify the relative fragility of a particular complex in electrospray. In this study, titration experiments were carried out at a cone voltage where the fraction of complex observed was not exquisitely sensitive to voltage.

4.6.2 VC50, affinity and enthalpy

Studies have suggested there might be a correlation between VC50 and affinity for a series of chemically similar ligands (Daniel *et al.*, 2002; Woods and Ferre, 2005). In this study VC50 appears to be a poor indicator of affinity in solution. However, there was correlation between VC50 and the enthalpic contribution to the free energy of binding (ΔH_{ITC}). There was also correlation with the number of hydrogen bonds between protein and ligand predicted by x-ray crystallography. It is worth noting that for the ligands selected with a similar affinity, if there is a correlation between VC50 and ΔH there will also be a correlation with $T\Delta S$ as ΔG is similar.

The protein binding domains of the macrocyclic ligands FK506 (0.4 nM) and rapamycin (0.2 nM) are very similar. The enthalpic contribution to the free energy of binding for the FKBP12-FK506 complex is large, ΔH_{ITC} = -14.2 kcal·mol^{-1}. 4 hydrogen bonds are seen between FK506 and FKBP12 in the x-ray structure. For the rapamycin complex ΔH_{ITC} = -16.4 kcal·mol^{-1} and 5 hydrogen bonds are seen between protein and ligand. These complexes did not dissociate under moderate cone voltages; VC50 were 45.8 and 48.8 V respectively. The x-ray structure of HEWL in complex with NAG3 predicts 6 hydrogen bonds between protein and ligand; this would suggest a large enthalpic stabilisation. The large enthalpic contribution to binding was confirmed by ITC, ΔH_{ITC} = -12.49 kcal·mol^{-1}. In solution, the

affinity was only in the low micromolar range (8 μM), this is explained by an unfavourable entropic contribution to binding, $T\Delta S_{ITC}$ = -5.53 kcal·mol^{-1}. The complex proved to be robust in electrospray, VC50 = 47.5 V. The interaction between RNaseA and 2'-CMP showed a similar thermodynamic profile. The x-ray structure of RNaseA in complex with 2'-CMP suggested 7 hydrogen bonds between protein and ligand. Charge-charge interactions are observed, non-covalent interactions of this type exhibit interaction energies much greater than a typical hydrogen bond, and this could account for enhanced stability of this complex in MS. A large enthalpic contribution to binding is confirmed by ITC (ΔH_{ITC} = -16.92 kcal·mol^{-1}). In solution the affinity is in the low micromolar range (1.5 μM) due to an unfavourable entropic contribution, $T\Delta S_{ITC}$ = -8.69 kcal·mol^{-1}. This complex was particularly robust in electrospray, VC50 = 55.8 V. High stability for ribonuclease-nucleotide complexes was recently reported by Loo's group (Yin *et al.*, 2008). Woods and Ferre (2005) performed collision-induced dissociation on a complex with an arginine-phosphate charge-charge electrostatic interaction using collision gas in a cell (ESI-MS/MS). Their complex consisted of a basic epitope containing adjacent arginine residues and an acidic epitope containing a phosphorylated serine. The authors describe how collision energies similar to covalent complexes were required for dissociation, they refer to this as a "covalent like" stability (Woods and Ferre, 2005). One might postulate that protein-DNA/RNA interactions might be particularly suitable for investigation by MS. In contrast to the studies mentioned, the FKBP-SD1 complex was prone to dissociation in MS but showed similar low micromolar affinity in solution (6.8 μM). The modelling studies predicted a lower enthalpic contribution to binding; only one hydrogen bond was predicted and there was a reduction in buried surface area compared to that seen in the x-ray structures of the more massive ligands in complex with FKBP12. This is the only complex in this study that showed a positive ΔS in ITC; ΔH_{ITC} = -6.09 kcal·mol^{-1}; $T\Delta S_{ITC}$ = 0.96 kcal·mol^{-1}. Additional evidence of a hydrophobic contribution to binding is provided by surface plasmon resonance. SPR was carried over a range of temperatures (6.4.5); non-linear van't Hoff analysis suggested a value of -0.71±0.14 kcal·K^{-1}·mol^{-1} for ΔC_p. A negative value for ΔC_p is indicative of a system where hydrophobic surface is removed from solvent into a nonpolar environment (Spolar *et al.*, 1989).

Evidence from the HEWL-NAG3 and RNaseA-2'-CMP systems support the hypothesis; complexes with a significant enthalpic contribution to the free energy of binding from electrostatic interactions survive electrospray better, is confirmed.

5 Chapter 5 The Spectral Ghost

– Influence of collision induced dissociation on charge-state distribution

5.1 Introduction

For the compounds tested in this study, it was not possible to see protein flying in complex if the ligand had a solution K_d > 160 μM. However, it was apparent that there were subtle changes to the charge-state distribution of the protein. In particular, the 8+ to 7+ intensity ratio appeared to reduce with an increasing concentration of ligand for those compounds known to bind in solution. We were interested if this effect were the result of complex dissociating under the electrospray conditions or if it were purely a consequence of increasing concentrations of ligand in the analyte solution. We considered the origin of charge on protein in electrospray and speculated if an extension to the charge residue model (CRM) proposed by Dole could explain the charge-state shifts in terms of the dissociation of the protein complex. The CRM proposes that drops contract by the evaporation of solvent until they reach their Rayleigh stability limit (Rayleigh, 1882) and undergo Coulombic fission (Dole *et al.*, 1968).

5.2 The origin of charge on multiply-protonated native FKBP12

In this study FKBP12 (5-20 μM) was sprayed from an aqueous solution of 10 mM ammonium acetate. The ratio of protein molecules to ammonium ions is in the order of 1:1000. It is thought that macromolecular charging is, in the most part, due to reaction of ammonium ions with basic residues on the surface of globular proteins (Felitsyn *et al.*, 2002; Touboul *et al.*, 2008a). The reaction of an ammonium ion with an amine leading to a proton bridged adduct is shown in equation 14. This process occurs in solution.

$$(\mathrm{FKBP12} \sim \mathrm{NH_2})^{Z+} + \mathrm{NH_4^+} \rightarrow (\mathrm{FKBP12} \sim \mathrm{NH_2} - \mathrm{H} - \mathrm{NH_3})^{(Z+1)+} \quad \textbf{Equation 14}$$

Where Z^+ is the charge-state of the protein and the $-NH_2$ group is part of a basic amino acid.

The complete proton transfer is thought to occur during dissociation of the adducted molecule during desolvation, equation 15.

$$(\mathrm{FKBP12} \sim \mathrm{NH_2} - \mathrm{H} - \mathrm{NH_3})^{(Z+1)+} \rightarrow (\mathrm{FKBP12} \sim \mathrm{NH_3})^{(Z+1)+} + \mathrm{NH_3} \quad \textbf{Equation 15}$$

However, it is also possible that a charged ammonium ion, NH_4^+, can dissociate leaving a protein ion with no additional charge, equation 16.

$$(\mathrm{FKBP12} \sim \mathrm{NH_2} - \mathrm{H} - \mathrm{NH_3})^{(Z+1)+} \rightarrow (\mathrm{FBP12} \sim \mathrm{NH_2})^{Z+} + \mathrm{NH_4}^+ \qquad \textbf{Equation 16}$$

Therefore, the final charge-state is not only a function of tertiary structure of the protein (the number of exposed basic residues) but also of the protein's ability to hold onto protons during dissociation of an adducted species.

The intrinsic gas-phase basicity of base B is defined as the free energy change ΔG^o on the loss of a proton, equation 17.

$$\mathrm{BH}^+ \rightarrow \mathrm{B} + \mathrm{H}^+ \qquad \textbf{Equation 17}$$

When considering the likelihood of an amino-acid chain becoming protonated, by the mechanism described in equations 14 and 15, one must consider the gas-phase basicity of the side chain. This can be thought of as the gas-phase equivalent of the pKa of a side chain in solution. Felitsyn *et al.*, (2002) considered the gas-phase basicities of basic residues on the surface of proteins and compared them to that of ammonia. The authors argue that it is possible for an NH_4^+ ion to protonate basic residues because the gas phase basicity of ammonia is less that that of arginine, histidine, lysine and to a lesser extent tryptophan and proline.

The usual assumption in ESI-MS is that a protein retains a correctly folded tertiary structure if the electrospray spectrum exhibits a narrow range of charge-states. The charge-state distribution for FKBP12 under standard ESI-MS conditions is dominated by the 8+ charge-state (Figure 2-8). In an acid solution (pH 2.5), FKBP12 denatures, more basic amino acids are exposed to solvent, and the electrospray spectrum shifts to a distribution centered on the 16+ charge-state (Figure 2-8).

The above argument possibly has relevance for the dissociation of protein-ligand complexes during the desolvation stages of ESI-MS. Is it possible for a ligand, of sufficient gas-phase basicity, to leave with a proton? This would have an effect of shifting the charge-state distribution to lower m/z. If this process occurs it should be possible to detect binding in

solution when no protein-ligand complex is observed as an ion in the ESI-MS spectrum due to dissociation.

5.3 Estimating the affinity of a ligand from a charge-state shift

Examining the charge-state distribution from titration experiments under mildly dissociating conditions revealed a charge shift with increasing concentration of ligand. Figure 5-1 (A) shows the ESI-MS spectrum of free FKBP12. Figure 5-1 (B) and (C) show the same protein solution with increasing concentrations of the 0.4 nM inhibitor FK506. The cartoon "blue T" represents the ligand and the "red sphere" FKBP12 in the FKBP12-FK506 complex flying as an 8+ charge-state (8+ molecular ion). Interestingly, these spectra showed the 7+ charge-state became the dominant charge-state for the un-complexed protein as the concentration of ligand increased.

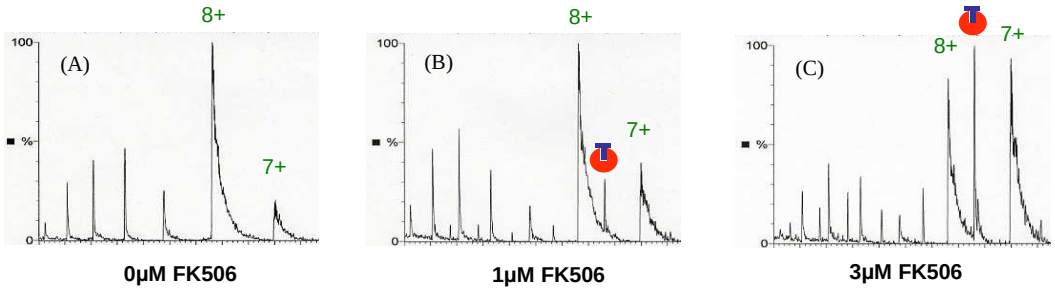


Figure 5-1 Changes to the charge-state distribution as a function of ligand concentration.

Analytes were dissolved in 10 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). **(A)** 5 µM FKBP12, 0 µM FK506 (B) 5 µM FKBP12, 1 µM FK506 (C) 5 µM FKBP12, 3 µM FK506. The peak for the 8+ complex is labelled with a blue T in a red sphere. Data is an average of 20 scans acquired on a Micromass ZMD instrument (Waters).

One hypothesis to explain the above spectra is that the loss of a protonated ligand during collision induced dissociation can reduce the charge-state of the protein. Hence, the 7+ signal is made up of macromolecular ions that initially charged at 7+ plus those that were in complex with FK506 (charged 8+) and then dissociated; FK506 leaving with a charge of 1+. FK506 is seen to fly as a $[FK506+H]^+$ ion in the electrospray spectrum. This process can be described by equation 18.

$$[FKBP12 \sim L]^{Z+} \rightarrow [FKBP12]^{(Z-1)+} + L^+ \qquad \textbf{Equation 18}$$

Where: L represents the ligand and Z+ the charge on the molecular ion prior to dissociation of the complex.

To investigate a correlation between the charge-state shift and the fraction of protein that had been in complex with ligand in solution, a metric was required that could follow changes to the CS spectrum of the protein. The metric chosen was the ratio of the intensity of the 8+ CS of the free protein ion divided by the 7+ CS of the free protein ion.

$$\frac{\text{Intensity of } 8+\text{CS}_{(\text{freeprotein})}}{\text{Intensity of } 7+\text{CS}_{(\text{freeprotein})}} = \frac{I^{8+}}{I^{7+}}$$

Figure 5-2 shows data from a series of titration experiments. Ligands were a selection from the test set (3.7.1). The x-axis shows an increasing total concentration of ligand. The y-axis on the left hand side shows the ratio I^{8+}/I^{7+} and the right hand axis the concentration of the protein ligand complex, [PL], calculated from the ratio of complexed to un-complexed protein (this method is described in section 3.6.3). In Figure 5-2 (A) and (B) [PL] was calculated from the spectrum. For Figure 5-2 (C) – (F) no complex was observed and [PL] is calculated from solution K_d (equation 12). To distinguish calculated and experimental [PL]; the experimental values are plotted in blue (solid line) and the calculated values in red (dashed line). The solid blue and dashed red lines are a least square fit to equation 12 using KaleidaGraph version 4.03 (Synergy Software). The solid green line is the fit of the I^{8+}/I^{7+} data to a logarithmic function as an aid to visualisation.

Figure 5-2 shows that there was a correlation between the charge-state shift, as monitored by the I^{8+}/I^{7+} ratio, and the fraction of protein in complex in solution. Alternative assays have detected complexes between the SD series and FKBP12 in solution (see section 6.4.6 (Stebbins *et al.*, 2007)). However these complexes with FKBP12 appeared not to be robust enough to survive ESI-MS. The dissociation of the complex could be described as leaving a ghost on the charge-state distribution (the spectral ghost of the title).

In Figure 5-3 I^{8+}/I^{7+} was plotted against fraction of protein that was in a complex with ligand in the solution phase. These results raise the intriguing possibility that it might be possible to estimate the affinity of a ligand by monitoring the charge-state shift under conditions that dissociate complex. This would extend the sensitivity of our instruments to beyond 160 μM. Detecting binding from the free protein spectrum has the additional advantage of not requiring the ligand mass to locate the signal from the protein in complex. This would mean that signal to noise might be less of an issue and acquisition times could be reduced.

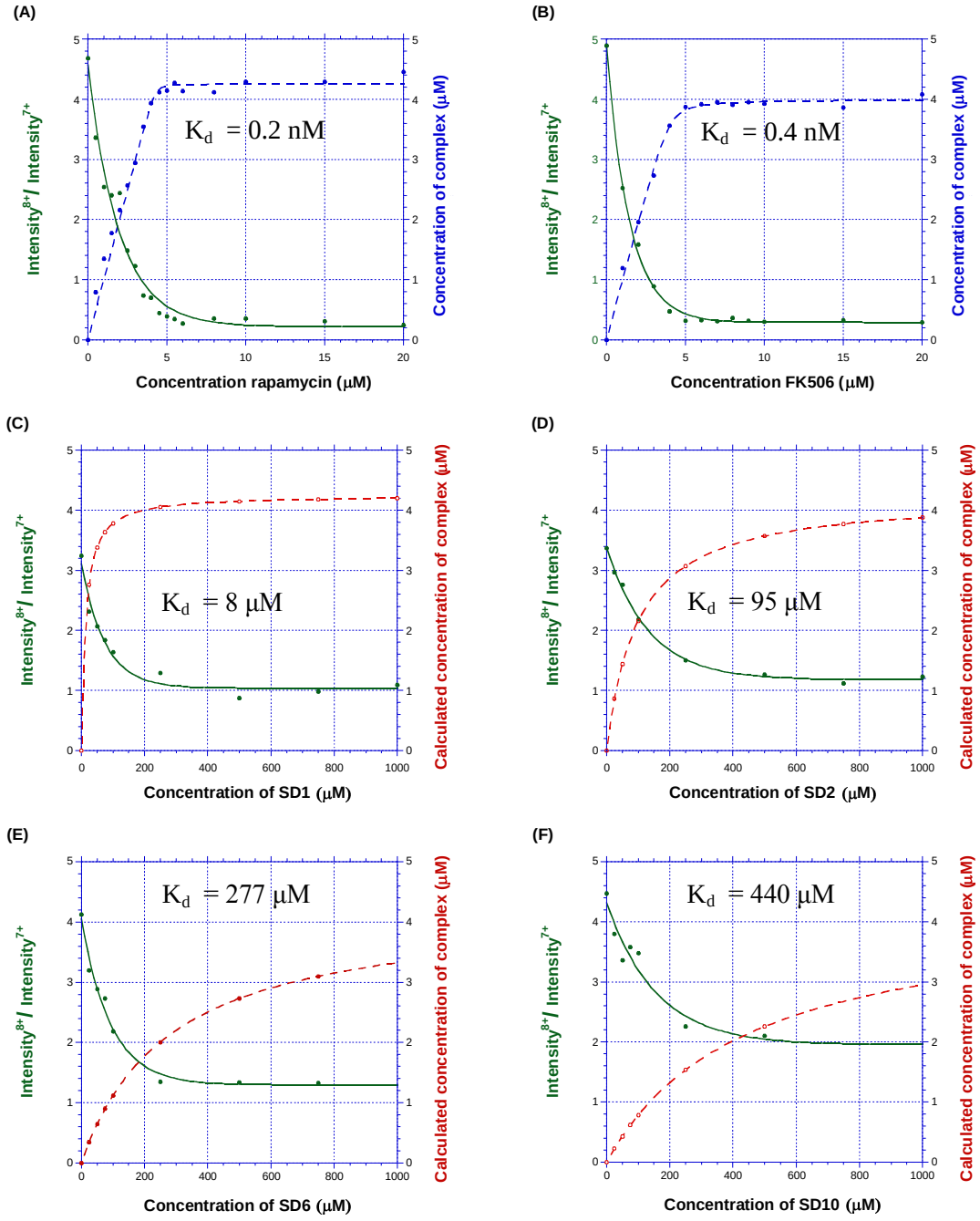


Figure 5-2 Charge-state shifts correlates with concentration of complex

Analytes were dissolved in 10 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). (A) 5 µM FKBP12, 0-20 µM rapamycin (0.2 nM) (dashed blue). (B) 5 µM FKBP12, 0-20 µM FK506 (0.4 nM). (dashed blue) (C) 5 µM FKBP12, 0-1000 µM SD1 (8 µM) (dashed red). (D) 5 µM FKBP12, 0-1000 µM SD2 (95 µM) (dashed red). (E) 5 µM FKBP12, 0-1000 µM SD6 (277 µM) (dashed red). (F) 5 µM FKBP12, 0-1000 µM SD10 (440 µM) (dashed red). Data is an average of 20 scans acquired on a Micromass ZMD instrument (Waters).

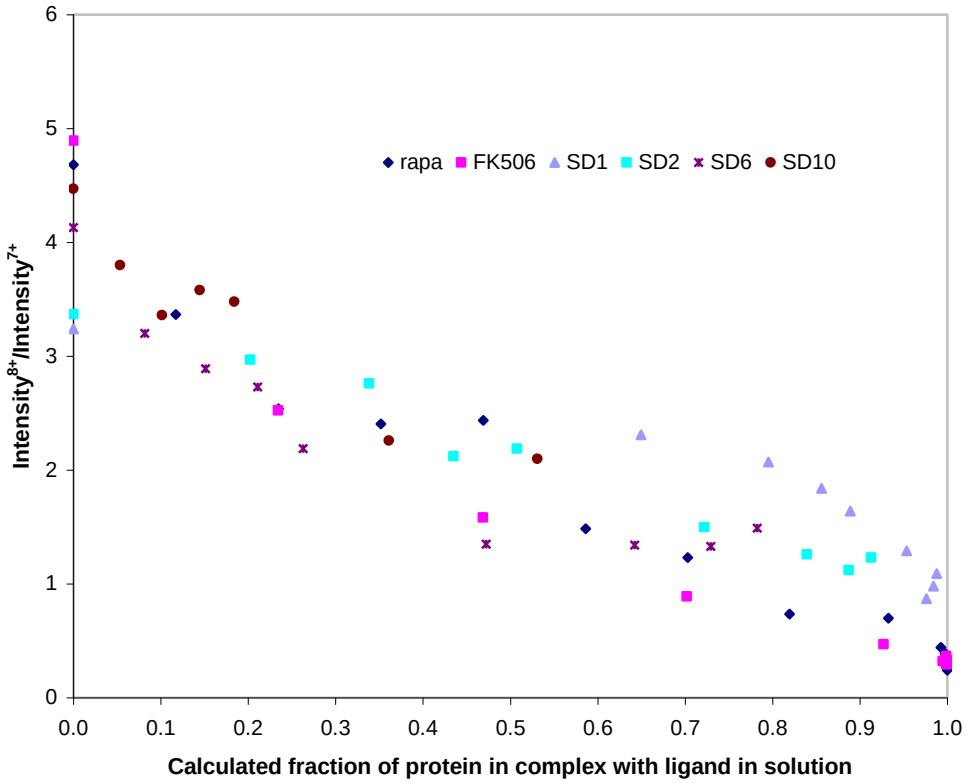


Figure 5-3 Charge-state shifts correlate with fraction of protein in complex with ligand in solution.

5 µM FKBP12. Analytes were dissolved in 10 mM ammonium acetate, pH 6.8, 10 % methanol (v/v) acquired on a Micromass ZMD instrument (Waters). The fraction of the protein in complex with the ligand was calculated from the concentration of the protein, the concentration of the ligand and the equilibrium dissociation constant using equation 12. Structures for each of the ligands are shown in Table 3-2 Ligand test set.

5.4 Collision induced dissociation and charge-state shift

To further investigate charge-state shifts and the ghosting effect; the charge-state shift of FKBP12 was examined under conditions of increasing collision induced dissociation. This was achieved by changing the cone voltage from 10 to 100 V. Dissociation caused by increasing cone voltage occurs in the late stages of the instrument after desolvation of the majority of adducted species and prior to mass analysis.

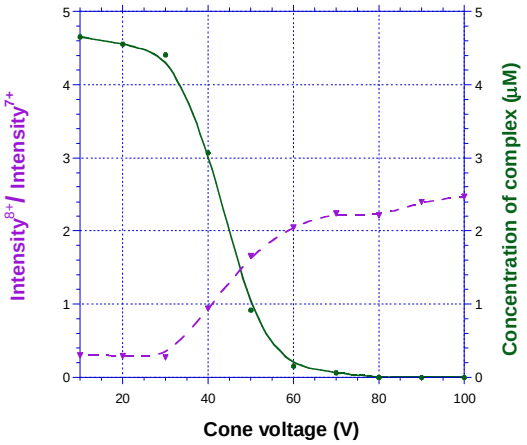


Figure 5-4 Charge-state shifts is a function of complex dissociation.

Analytes were dissolved in 10 mM ammonium acetate, pH 6.8; 10 % methanol (v/v); 5 µM FKBP12; 10 µM FK506 (0.4 nM). Data is an average of 20 scans acquired on a Micromass ZMD instrument (Waters).

The results shown in Figure 5-3 show that there is a correlation between the fraction of the protein in complex with the ligand and the charge-state distribution of the protein, as measured by the metric I^{8+}/I^{7+}. Careful examination of a series of spectra produced by increasing cone voltage, as seen in Figure 3-6 (A), would suggest that loss of the ligand from a molecular ion charged 8+ can go through two mechanisms. At lower cone voltages (10 - 40 V) the spectra suggested that the dominant mechanism is the dissociation of the complex molecular ion (8+ CS) through loss of a protonated ligand (L+) to give a 7+ free protein ion as described by equation 18. At low cone V the 7+ and 8+ CS for the free protein were minor components within the spectrum (dominated by the 8+ CS for the FKBP12-FK506 complex). As the cone V increased the 7+ CS for the free protein increased at a faster rate than the 8+ charge-state for the free protein. This remained the case up to 30 V when the fraction of the 8+ CS for free protein increased to become the dominant species. The

increase in the 8+ CS for the free protein is explained by the loss of a neutral ligand ion from the complex.

5.5 Estimating the affinity of a ligand from a charge-state shift under dissociating conditions

To further test the hypothesis that the dissociation of the ligand is in the latter stages of the electrospray process caused a charge-state shift, a titration experiment was performed between FKBP12 and FK506 under fully dissociating conditions, Figure 5-5. The titration experiment with FKBP12 and FK506 was performed with a cone voltage of 60V.

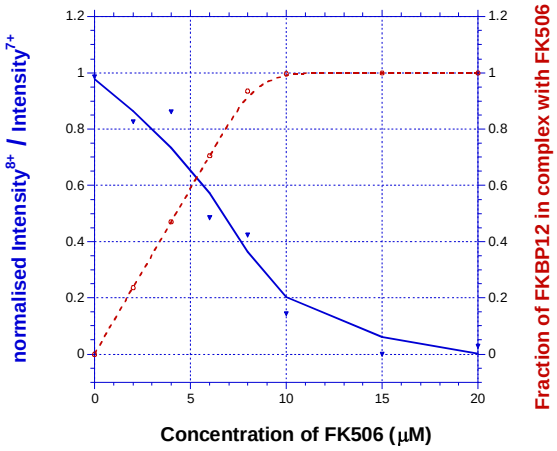


Figure 5-5 Charge-state shifts correlate with fraction of protein in complex with ligand under dissociating conditions.

Analytes were dissolved in 10 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). 5 μM FKBP12, 0-20 μM FK506 (0.4 nM). Dashed red line is the calculated fraction of FK506 in complex with FKBP12 scaled between 0 and 1. Data is an average of 20 scans acquired on a Micromass ZMD instrument (Waters).

The results represented in Figure 5-5 showed that it was possible to estimate the fraction of protein in complex with ligand in solution under conditions of complete dissociation by considering the charge-state shift as measured by the metric I^{8+}/I^{7+}.

5.6 Charge-state shift as a function of complex dissociation in other protein systems

A change in charge-state distributions on dissociation of the complex are also observed for the HEWL-NAG3 and RNaseA- 2'-CMP complexes (Figure 5-6 and Figure 5-7 respectively).

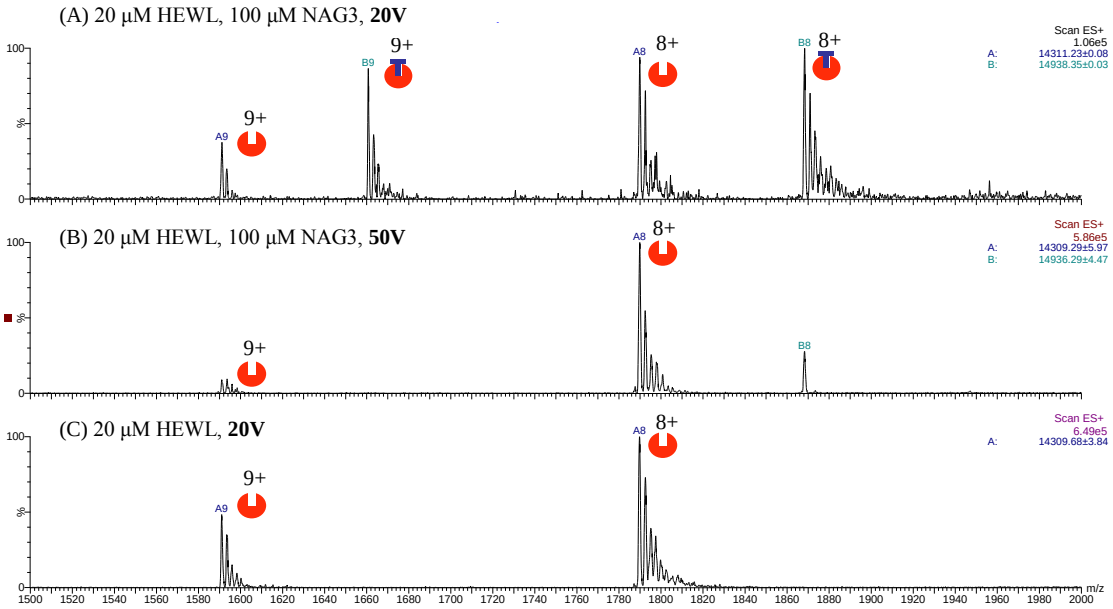


Figure 5-6 Dissociating the HEWL-NAG3 complex

Analytes were dissolved in 20 mM ammonium acetate, pH 6.8 at the concentrations detailed on the spectra. Spectra are annotated with cone voltage. Complex ions and free protein ions are labelled with icons to aid identification (blue "T" represents the ligand and the "red sphere" the protein). The free protein ion series are labelled with A9 and A8, representing the 9+ and 8+ charge-states. The complex ion series are labelled with B and the charge-state.

There is evidence from Figure 5-6 (A) and (B) that at a cone voltage of 50 V the 9+ complex ion (HEWL-NAG3) has dissociated into an 8+ free protein ion in the dominant dissociation pathway, as described by equation 18.

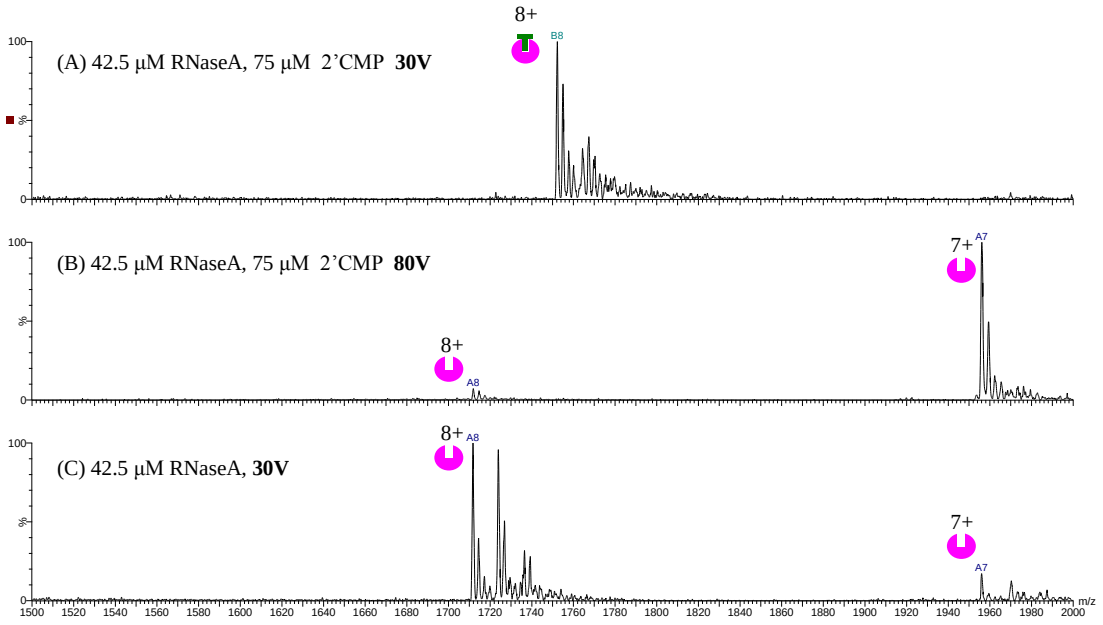


Figure 5-7 Dissociating the RNaseA-2'CMP complex.

Analytes were dissolve in 20 mM ammonium acetate, pH 6.8 at concentrations detailed on the spectra. Spectra are annotated with cone voltage. Complex ions and free protein ions are labelled with icons to aid identification (green "T" represents the ligand and the "magenta sphere" the protein). The free protein ion series are labelled with A8 and A7, representing the 8+ and 7+ charge-states. The complex ion series are labelled with B and the charge-state.

There is evidence from 5.7 (A) and (B) that when the cone voltage is increased from 30 V to 80 V the 8+ RNaseA- 2'-CMP complex ion dissociated into a 7+ free protein ion and 2'CMP (not seen in the illustrated mass range). The mechanism for this could be described by equation 18.

5.7 How consistent is desolvation?

The desolvation of the protein has implications for the quantification of affinity for both the direct measurement of peak intensities and indirectly from a charge-state shift. Under conditions necessary to preserve the complex in electrospray each charge-state has an adduct tail indicative of incomplete desolvation. The minimum m/z value for a given charge-state was defined as the base peak; this represents the ion that has undergone the most complete desolvation, Figure 5-8 (A). The adduct tail corresponds to ions with an increased m/z due to non-specific binding of species present in the analyte that have not been removed by desolvation. Examples of non-specific binders are H_2O and NH_4^+. The Micromass ZMD and ZQ instruments (Waters) were found to have reasonably consistent desolvation with relatively small adduct tails for each charge-state. The LCQ^{TM}DECA (ThermoQuest) was the more sensitive instrument in terms of retaining low affinity ligands in complex with protein. However, desolvation appeared to be less consistent from visual inspection of the spectra. The titration data also showed more variability than that obtained with the Micromass ZMD/ZQ. Data acquisition with this instrument was approximately 3 times longer than with the Waters instruments.

It would be convenient to use the base peak as an approximation for the whole charge-state envelope as charge-state envelope overlap need not then be deconvoluted. The variability of desolvation and I^{8+}/I^{7+} charge-state ratio of the free protein were investigated for 7 samples of FKBP12 flying under optimized conditions on the LCQ^{TM}DECA. The intensity of the charge-state was calculated from the summation of ion intensities in the charge-state envelope. The study investigated how much of the adduct tail needed to be included in the calculation to give an I^{8+}/I^{7+} charge-state ratio with low variance. The intensity of the charge-state was calculated for Δm between 0 and 900 Da, equation 19.

$$\text{Intensity}^{ChargeState} = \sum_{i=m}^{i=m+\Delta m} I_i \qquad \textbf{Equation 19}$$

Where: $I^{ChargeState}$ is the total intensity for a specific charge-state envelope, m is the base peak for the charge-state and Δm is the mass equivalent above the base peak mass that is included in the intensity summation.

Figure 5-8 (B) shows a plot of the difference in I^{8+}/I^{7+} between successive Δm, $(I^{8+}/I^{7+})_{\Delta m+1}$- $(I^{8+}/I^{7+})_{\Delta m}$, plotted against Δm, where Δm has been rounded to the nearest integer. Rounding Δm to the nearest integer was necessary when combining data from more than one experiment. It can be seen that $(I^{8+}/I^{7+})_{\Delta m+1}$- $(I^{8+}/I^{7+})_{\Delta m}$ becomes asymptotic when Δm > 200 Da. Figure 5-8 (C) shows how the mean I^{8+}/I^{7+} (N=7) reaches a steady value with reduced variance when Δm > 200 Da. Considering a model adduct with a mass of 20 Da one can deduce that there is most variability in desolvation for the final 10 non-specifically bound species under the experimental operating parameters. It is interesting to note that in the rapamycin and FK506 titration data the charge-state envelope corresponding to the protein in complex with the ligand shows better desolvation than the free protein. This would be consistent with hypothesis that ligand occupies the active site of FKBP12 and solvent molecules have lower affinity out-with the active site. There are 3 conserved water molecules in the apo structure of FKBP12 (2PPN.pdb, Figure 1-8). Displacement of ordered water on ligand binding has been calculated to give a favourable entropic contribution the free energy of binding (Connelly *et al.*, 1993).

Figure 5-8 (D) shows titration data plotted using the base peak (dotted green) and values calculated using a summation of Δm = 200 Da (solid blue). Visual inspection of the data shows a lower value for the saturation concentration of complex for the data including Δm = 200 Da (solid blue) in the summation. The conclusion was that there were no appreciable benefits in including the adduct tail in the estimation of ligand affinity.

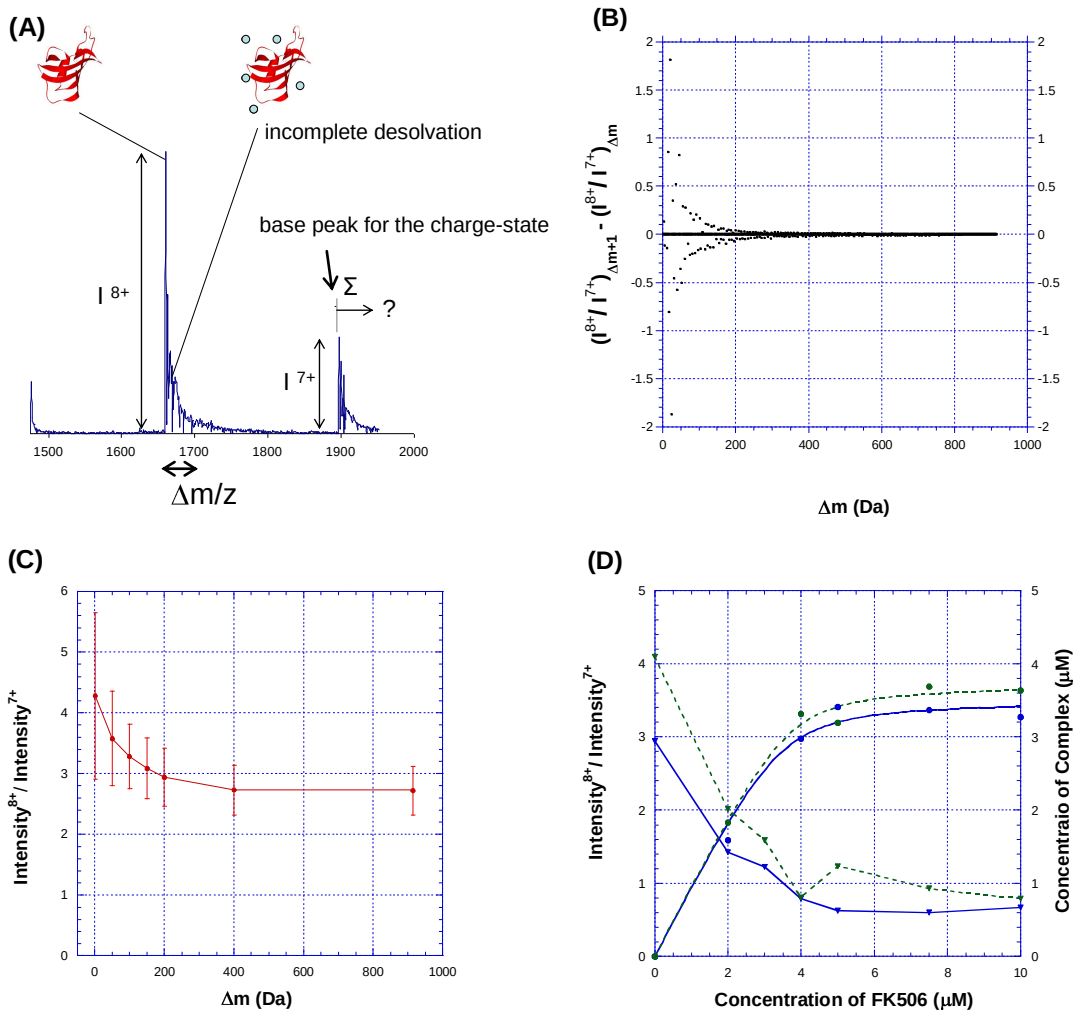


Figure 5-8 Does inclusion of the adduct tail lead to better agreement with solution phase affinity data?

(A) Annotated 8+ and 7+ charge-states of FKBP12 showing adduct tails. (B) A typical plot of the difference in I_{8+}/I_{7+} as a function of Δm, $(I_{8+}/I_{7+})_{\Delta m+1}$- $(I_{8+}/I_{7+})_{\Delta m}$, plotted against Δm, where Δm has been rounded to the nearest integer. (C) Mean I_{8+}/I_{7+} as a function of Δm, error bars are the standard deviations for N=7. Data (A), (B), (C) 4 μM FKBP12, in 10 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). (D) Titration experiment, points are a mean for N=3. 4 μM FKBP12, 0-10 μM FK506, in 10 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). Each scan was an average of 300 scans acquired on a LCQ^{TM}DECA (ThermoQuest) instrument; capillary voltage 3.52 KV, cone voltage 25 V and capillary temperature 60 ^{o}C.

5.8 Discussion and conclusions

The use of charge-state shifts to assess the affinity of weak ligands that dissociate under electrospray conditions is an intriguing concept. There is preliminary data to support the method having enhanced sensitivity compared to detecting binding from the presence of a complex ion in a standard ESI-MS screening experiment. This technique may have particular application when screening small fragment-like molecules which have affinity in the micromolar range but become potential lead compounds after fragment linking or chemical modification.

However, measuring ligand binding by this method is still at a preliminary stage of investigation. It would seem reasonable to suggest that the mechanism of charge division is very likely to be ligand dependent in any given protein system. The gas phase basicity of the ligand must have an influence on the dissociation pathway, as must the nature of the non-covalent interactions between ligand and protein. It is however useful to observe that a charge-state shift might indicate the dissociation of a complex in electrospray and that further investigation is required. If time had allowed we would have investigated carbonic anhydrase, many small soluble inhibitors are available over a range of K_d (Krishnamurthy *et al.*, 2008). For a larger data set it would be interesting to relate charge state-shifts to the relative basicity of hydrogen bond acceptors on ligands. Laurence *et al*. have collated a database of constants measuring the hydrogen-bond basicity of organic species towards a reference hydrogen-bond donor (Laurence *et al.*, 2009).

The conclusion reached in the experiments investigating “adduct tails” was that there were no appreciable benefits in including the adduct tail in the estimation of ligand affinity by summation.

6 Chapter 6 Biophysical and Biochemical Ligand Binding Studies of FKBP12

6.1 Introduction

This study employed a variety of biophysical and biochemical techniques to characterise non-covalent complexes of FKBP12. One of the principle aims of the project was to explore medium throughput biophysical screening strategies for testing hits from virtual screening programs. Electro-spray ionisation mass spectrometry (ESI-MS) is discussed in detail in chapters 3, 4 and 5. It was important to compare results from ESI-MS with other experimental approaches. Choice of biophysical method required balancing throughput, sensitivity and the richness of thermodynamic data. Compounds from the ligand test set were used to compare different methods (section 3.7 describes the selection of the test set).

It is useful to make a distinction between a screening experiment in which a large number of ligands are tested at a single concentration and the thermodynamic characterisation of a specific complex where some form of titration is employed. The estimation of affinity at a single ligand concentration is inherently less accurate than a titration method. However, the increased throughput outweighs this disadvantage when used as the first experimental step in the drug design process. Some methods are more suitable for screening at a single ligand concentration because of a low level of variance between repeats. Other techniques have the advantage of a 96-well plate format or in the degree of automation of data collection. Some methods are particularly prone to false positives. Coloured ligands can be a problem for assays that are dependent on a colour change such as the PPIase assay. Similarly the potential for ligand fluorescence must be considered.

The threshold affinity for the detection of a complex, at a specific protein/ligand screening ratio, is of significance. Fragment-based screening approaches are of current interest in drug design. Fragment-like ligands are generally defined as those with a MW $\leq$ 350 Da (Carr *et al.*, 2005). This tends to limit the number of non-covalent interactions accessible for a specific fragment when compared to that of a larger drug. Equilibrium dissociation constants (K_d) of hits from fragment screens are often between 50 and 150 μM (Congreve *et al.*, 2008). These ligands can be challenging to detect by some biophysical techniques, particularly if they are relatively insoluble.

6.2 General screening considerations

It is self-evident that the protein/ligand screening ratio determines the minimum affinity of ligand detected and the range of affinity over which there is good discrimination between ligands of close affinity. For each assay, there is minimum protein concentration required to give sufficient response; for FKBP12 this is around 10 μM for the biophysical techniques used in this study. This concentration of protein necessitates a ligand concentration of between 100 and 250 μM to detect micromolar affinity ligands in a screen. To illustrate this it is useful to consider a weak inhibitor with K_d = 250 μM. At a screening concentration of 100 μM 28 % of the protein would be in complex with the ligand and at 250 μM 50 %, this not a linear relationship. However, for the thermal denaturation assay a ligand screening concentration of 250 μM equates to a change in melting temperature of less that 1 ^{o}C for this ligand; small compared to the standard deviation for three repeats which is in the order of 0.5 ^{o}C. There are often solubility issues for ligands at a concentration of 250 μM in an aqueous buffer. Compounds from screening libraries are routinely dissolved in DMSO or DMSO is added to an analyte solution to aid solubility. Screening compounds against FKBP12 is complicated by the fact that DMSO binds to the protein with an affinity in the low millimolar range (Burkhard *et al.*, 2000). 5 % DMSO is equivalent to a 0.7 M solution; a close to saturating concentration if no competing ligand is present. Introducing a low % of DMSO into solution can lead to significant competition for the active site with a weakly binding screening compound at the usual protein/ligand screening ratios.

Table 6-1 illustrates the effect of 0 – 5 % DMSO in 3 hypothetical experiments with a high affinity ligand (K_d = 0.7 μM), a medium affinity ligand (K_d = 7 μM) and a low affinity ligand (K_d = 70 μM). The model assumes 10 μM FKBP12, 250 μM screening ligand and a K_d = 30 mM for DMSO. The ligand simulation assumes the law of mass action and that the DMSO and the screening ligand compete for the same binding site. A solution of the algebraic equation describing complex formation as a function of total ligand concentration was arrived at using the program Berkley Madonna (V.8.3.11, Macey and Oster). The concentrations of the protein complexes and free protein were calculated at equilibrium. For the 0.7 μM ligand a low % of DMSO has little effect on the proportion of protein seen in complex with the screening ligand. However, there is significant competition for the binding site for the ligand with a K_d = 70 μM. This result has implications for including DMSO, or any solvent that binds both specifically or non-specifically, in a screening assay. There is a real possibility that low affinity ligands/fragments might appear as false negatives.

(A) Ligand K_d = 0.7 μM (high affinity),		Total [FKBP12] = 10 μM ,		Total [Ligand] = 250 μM
% DMSO	***Total [DMSO] (mM)***	***Eq. [FKBP-L] (μM)***	***Eq. [FKBP12-DMSO] (μM)***	***Eq. [FKBP12] (μM)***
0	0	9.97	0	0.03
0.1	14	9.96	0.01	0.03
1	141	9.84	0.13	0.03
2	282	9.72	0.25	0.03
5	715	9.36	0.61	0.03

(B) Ligand K_d = 7 μM (medium affinity),		Total [FKBP12] = 10 μM ,		Total [Ligand] = 250 μM
% DMSO	***Total [DMSO] (mM)***	***Eq. [FKBP-L] (μM)***	***Eq. [FKBP12-DMSO] (μM)***	***Eq. [FKBP12] (μM)***
0	0	9.72	0	0.28
0.1	14	9.60	0.12	0.28
1	141	8.65	1.10	0.25
2	282	7.80	1.98	0.25
5	715	5.99	3.83	0.18

(C) Ligand K_d = 70 μM (low affinity),		Total [FKBP12] = 10 μM ,		Total [Ligand] = 250 μM
% DMSO	***Total [DMSO] (mM)***	***Eq. [FKBP-L] (μM)***	***Eq. [FKBP12-DMSO] (μM)***	***Eq. [FKBP12] (μM)***
0	0	7.76	0	2.24
0.1	14	7.07	0.9	2.03
1	141	3.94	4.94	1.12
2	282	2.65	6.60	0.75
5	715	1.32	8.31	0.37

Table 6-1 DMSO as a competing ligand

The concentrations of the species at equilibrium were calculated for three hypothetical ligands: (A) K_d = 0.7 μM, (B) K_d = 7 μM and (C) K_d =70 μM ; [ligand] = 250 μM, [FKBP12] = 10 μM and DMSO between 0 and 5 %. The program Berkeley Madonna (V.8.3.11) was used to solve the equations for two competing ligands (Appendix 1). The concentrations of the protein complexes and free protein were calculated at equilibrium (Eq.). Including DMSO in a screening experiment with a tight binding ligand, K_d = 0.7 μM (A) had little effect on the proportion of protein in complex with the ligand. For a weaker inhibiter, K_d = 70 μM (C), there was significant competition at 1 % DMSO.

It is also interesting to note that for the 70 μM ligand although there is significant competition with DMSO at < 5 %; the ratio of [FKBP-ligand]/[FKBP] remains remarkably consistent (FKBP12-DMSO complex is the dominant species in solution).

The screening ratio was biased for the detection of low affinity ligands. This reduced discrimination between nanomolar affinity ligands. In the unlikely eventuality that a large number of tight binding ligands were discovered in a single screen, more detailed thermodynamic characterisation would discriminate between them.

6.3 Biochemical and biophysical methods

6.3.1 Peptidyl-prolyl isomerase assay

In this study the K_d for inhibitors of FKBP12 was determined from the competitive inhibition of the isomerisation of the substrate Succ-Ala-Leu-Pro-Phe-*p*-nitroaniline.

An assay was developed by Fischer to measure peptidyl-prolyl isomerase (PPIase) activity in immunophilins (Fischer *et al.*, 1984a; Fischer *et al.*, 1984b). This assay relies on the conformationally selective hydrolysis of substrates of the form Xaa-Pro-Phe-*p*-nitroaniline by α-chymotrypsin. Hydrolysis of the C-terminal *p*-nitroaniline ester bond occurs for Xaa-Pro is in the *trans* conformation. FKBP12 catalyses the conversion between the *cis* to *trans* conformation of the peptidyl-prolyl bond in the substrate. FKBP12 shows substrate specificity for peptides with a hydrophobic, bulky residue next to the proline (Harrison and Stein, 1990). The progress of the ester hydrolysis may be followed spechrophotometrically by measuring the absorbance of the bright yellow *p*-nitroaniline product (λ_{abs} = 400 nm). Although theoretically sound, the reaction was found to be hard to quantify because of uncertainty in the initial ratios of *cis*/*trans* isomers. The *trans* form of the substrate was hydrolysed very rapidly by α-chymotrypsin in the so called "burst" reaction. Moreover, there was a high thermal turnover between *cis* to *trans* conformers at room temperature. The assay was improved by Kofron *et al.*, (1991) who introduced a new solvent for the substrate that shifted the conformational equilibrium in favour of the *cis* isomer (Kofron *et al.*, 1991).

A possible disadvantage of this assay is the dependence on a colour change; this can require careful use of controls for coloured or fluorescent compounds. In addition, the reaction occurs over a short time course. The *cis/trans* ratio of the substrate is reduced in the presence of water. The substrate must be kept as dry as possible prior to and during experimentation. The experimental method required considerable manual dexterity.

Materials and Methods

In this study Kofron's method was followed (Kofron *et al.*, 1991). A stock solution of 6 $mg^{-1} \cdot ml^{-1}$ α-chymotrypsin (CalBiotech) in 10 mM hydrochloric acid was prepared and kept on ice. Buffered FKBP12 was incubated with each inhibitor for at least 30 minutes at 0 ^{o}C. 100 μl of α-chymotrypsin was added to the protein/ inhibitor solution, gently mixed and very rapidly pipetted into a spectrometer cuvette containing 30 μl, 4 mM substrate (Bachem,UK) in 2,2,2-trifluoroethanol, 47 mM lithium chloride (placed in the instrument). The substrate

solution was pipetted immediately before the rapid pipetting step. This ensured that the substrate remained dry and *cis/trans* ratio was as high as possible. Final solution: 100 nM FKBP12; x μM ligand; 0.6 $mg^{-1} \cdot ml^{-1}$ α-chymotrypsin; 120 μM Succ-Ala-Leu-Pro-Phe-*p*-nitroaniline; 14 mM lithium chloride; 1 % methanol (v/v); 3 % 2,2,2-trifluoroethanol (v/v), PBS, pH 8; final volume 1 ml. The initial equilibrium ratio of *cis*/*trans* Xaa-Pro-Phe-p-nitroaniline was found to be ~ 1.5. This allowed a measurable time course for the isomerase reaction and a reduction in the initial burst reaction when α-chymotrypsin rapidly hydrolysed the *trans* conformer. Thermal turnover was reduced by conducting the assay at 4^{o}C (Peltier cooling of cuvette). It was not possible to cool below 4 ^{o}C due to problems with condensation on the cuvette during the course of the reaction. Absorbance at 400 nM was record every 0.01s for 120s on a Perkin-Elmer Lambda 20 spectrometer (Perkin- Elmer). The rate of the reaction was determined between 0.5 and 1.5 s and rates determined using UV Kinlab software (V1.0, Perkin- Elmer) and an extinction constant of $\varepsilon_{400nm} = 10{,}050\ M^{-1} \cdot cm^{-1}$. From 0-2 s the substrate was assumed to be in excess and the rate of isomerisation independent of the substrate concentration. This was seen as a linear portion of the graph of absorbance of product plotted against time. The reaction reached completion within 120 s.

Without the use of stop-flow equipment there was still considerable variance in the experimental data; each data point was repeated 9 times and the standard error calculated. The random errors in data; are believed to be due to variability in the start of data recording, incomplete mixing and bubbles in the cuvette because of over vigorous pipetting. The activity of α-chymotrypsin is pH dependent; maximum activity occurs at pH 8.0 (Blow *et al.*, 1969). Solution pH was confirmed for each test compound and a control was included to check for the possible effect of the inhibitor (no protein) on background turnover of the substrate. This control also accounted for coloured or fluorescent compounds. It was not possible to test some potential inhibitors due to their intense yellow colour.

Screening the test set by PPIase assay

The PPIase assay was used to screen members of the inhibitor test set to assess the threshold affinity of the assay for a specific compound screening concentration. The experiment was also designed to assess the variance in a reading and the suitability of the assay to rank order a series of compounds by affinity. A concentration of 300 μM was selected; most compounds were soluble to this point. Ligands were dissolved in methanol (no DMSO) and solvent concentrations kept constant in every experiment. The % inhibition was calculated at the screening concentration.

Determining the apparent equilibrium dissociation constant, K_i

The concentration of the inhibitor was varied from zero (0 % inhibition) to 100 % inhibition of the PPIase activity of FKBP12. 100 % inhibition was not attainable for lower affinity ligands of poor solubility and restricted the general applicability of the method.

The K_i, for an inhibitor was determined by the non-linear least squares fit of equation 20 to the rate of turn-over of the substrate, corrected for the background turnover at 4 ^oC, plotted against the concentration of an inhibitor (Morrison, 1969; Kofron *et al.*, 1991). Non-linear regression was performed using the program Kaleidagraph (V4.03 Synergy software).

$$V_i = \frac{V_o}{2\ P} \times \left\{ \ P\ -\ I\ \ -K_i\ + \sqrt{\ P\ -\ I\ -K_i\ ^2 + \ 4 \times\ P\ \times K_i} \right\} \qquad \textbf{Equation 20}$$

Where; V_i is the rate of the reaction at a specific concentration of inhibitor, V_o the rate of reaction with no inhibitor present, [P] is the concentration of FKBP12 and [I] the concentration of the inhibitor.

The equilibrium dissociation constant, K_d, for each inhibitor, was determined by correcting the K_i for competition with the substrate using equation 21 (Cheng and Prusoff, 1973).

$$K_d = \frac{K_i}{1 + \left(\frac{sub_{\ cis}}{K_m} \right)} \qquad \textbf{Equation 21}$$

Where: $[sub]_{cis}$ is the concentration of the *cis* isomer of the substrate and K_m is the Michaelis constant for substrate . K_m = 301 μM for the 6H-FKBP12 construct used in this study (Wear *et al.*, 2007b).

The concentration of the substrate was taken as the concentration of the *cis* isomer. This was calculated from the total concentration of the substrate corrected for the proportion of the substrate in the *trans* form. This was determined from the absorbance at close to zero, after the "burst" reaction, as a fraction of the total absorbance after completion of the reaction.

6.3.2 Isothermal titration calorimetry

A defining feature of isothermal titration calorimetry (ITC) is the direct measurement of the enthalpy of binding, ΔH, for a protein-ligand complex. ITC also provides information on stoichiometry, the free energy change, ΔG, and the entropic change, ΔS (Cooper and Johnson, 1994; Holdgate, 2001; Freyer and Lewis, 2008; Bouchemal, 2008). The K_d for the interaction can be calculated from ΔG using equation 2. It is possible to perform experiments at a range of temperatures, examine changes in the thermodynamic profile with temperature and calculate the difference in specific heat capacity, ΔC_p, for a complex.

In a typical ITC experiment the ligand is injected into a solution of protein and the heat change measured. The heat change comes from formation of the complex, the heat of dilution of the ligand and protein and a small effect due to mixing (the injection syringe constantly rotates in the reaction cell). It was necessary to perform controls and correct for dilution and mixing effects in ITC experiments, Table 6-2. The heat of dilution of the ligand was typically the largest correction necessary in this study. An exception to this rule, were the experiments using ammonium acetate as a buffer and injecting a concentrated solution of protein into a ligand solution. There was a significant heat of dilution for the protein in buffer.

ITC was principally used for secondary screening; the K_d for a given complex was estimated from primary screen prior to the ITC experiment. The purpose of the ITC experiments was to improve the estimation of K_d and to determine other thermodynamic parameters.

Materials and Methods

Concentrated solutions of ligand were added to a solution of protein in a series of small injections and the heat change for each injection measured by comparison to a reference cell containing the experimental buffer. The law of mass action and an estimated K_d were used to model the complex formed for each proposed injection (Microsoft Office Excel, 2003). Analyte concentrations and injection volumes were adjusted so that the initial injections had similar integrated heats of injection and that the binding sites of the protein sample were saturated for the final few injections. Sigmoid curves were achieved for a plot of integrated heats of dilution against the molar ratio of ligand/protein (after correction for heats of dilution/mixing) for the high affinity ligands. Integrated heat curves for lower affinity ligands show less steep sigmoid curves for the same analyte concentrations. Beyond the

obvious influence of affinity on the shape of the curve there were usually experimental constraints from ligand solubility and quantity of available protein. The test set was a useful tool to assess the discrimination between different ligands in experiments designed with a low c value. c = (concentration of macromolecule x stoichiometry)/K_d and is a useful dimensionless quantity used to assess the sensitivity of the ITC thermogram to subtle changes in K_d (Holdgate, 2001)). Increasing the c value has an obviously detrimental effect on protein requirement in an already protein hungry assay.

ITC experiments were carried out on a MicroCal auto-VPITC instrument (GE Healthcare). Reagents were dissolved in 50 mM sodium phosphate, pH 6.8; 50 mM sodium chloride. Ammonium acetate was initially tested as a buffer as it is used in ESI-MS experiments. An advantage of ammonium acetate is that the pKa does not vary much with temperature between 5 and 60 ^{o}C. However, it was found that the heats of dilution for the protein was large compared to enthalpy of ligand binding. The majority of the experiments were performed with a sodium phosphate buffer (Ferguson *et al.*, 1980). In a screening experiment 300 μM ligand (in the syringe) was injected into 15 μM FKBP12 (in the reaction cell). For insoluble ligands it was possible to inject a concentrated solution of protein into a ligand solution. Rapamycin and FK506 were tested by injecting solutions of 35 μM FKBP12 (syringe) into 1 μM ligand (reaction cell). The default experimental protocol was: 28 x 10 μl injections (0.5 μl 1st injection), 20 s injection duration, 180 s injection delay, syringe spin 240 rpm, reference power 10 μcal·s^{-1}, 25 ^{o}C.

The auto-VPITC allowed samples of protein and ligand to be stored at 4 ^{o}C prior to experimentation. Analyte solutions are degassed as part of the automated filling of the syringe and reaction cell. The reaction cell and syringe were automatically cleaned between experiments.

Control Experiments

Table 6-2 shows the control experiments necessary for each ligand examined. Each experiment was repeated in triplicate. DMSO and methanol are often chosen as the solvent for ligand stock solutions. It was important to match solvent concentration in the reaction cell, syringe and reference cell to reduce unnecessary solvent dependent effects on the heats of injection.

Data Analysis

Heats of reaction were determined by integration of heat of injection peaks and baseline corrected for heats of dilution. In most cases, it was possible to fit the heats of dilution to a linear model. This was then used to correct the data. This method meant that unnecessary noise was not added to the heat of formation of the complex. Integrated heat data was fit to a standard 1:1 binding model, after baseline correction, using the instrument software (Origin V4.0, MicroCal).

Cell $(2ml)_{total}$	***Syringe*** $(0.7ml)_{total}$	***Information from experiment***
protein	ligand	enthalpy of complex formation combined with heats of dilution/mixing
protein	buffer	heat of dilution of protein
buffer	ligand	heat of dilution of ligand
buffer	buffer	heat of mixing, possible buffer mismatches, poor cell rinsing

Table 6-2 Typical control experiments for ITC

Errors in analyte concentration

Data analysis in ITC experiments is very sensitive to errors in analyte concentrations. The deviation of the "chemical" stoichiometry, N, from 1 (or a known real number) is an indication of the systematic error in the experiment. These errors tend to be dominated by errors in the protein and ligand concentrations. It is worth noting that fixing N to 1 does not lead to improvements in the estimation of ΔH and ΔS due to the correlation of the parameters and must be avoided (Tellinghuisen, 2003). Determining the concentration of the protein is described in section 2.3.2. The concentration of rapamycin was determined from A_{295nm} using an extinction coefficient of 30, 500 $M^{-1} \cdot cm^{-1}$ (Wear *et al.*, 2007b).

6.3.3 **Differential scanning calorimetry**

The temperature induced unfolding (denaturation) of a protein can be investigated by differential scanning calorimetry (DSC). The heat energy required to raise the temperature of a protein sample is compared to that required for a similar protein free solution in a reference cell. Most proteins are only marginally stable under physiological conditions. Heating, a solution of a folded protein, increases the internal energy of the molecule and leads to disruption of non-covalent bonds holding the amino acid chain in the folded state (Jelesarov and Bosshard, 1999). Investigation of the folding and unfolding pathways of FKBP12 showed that wild-type protein folds and unfolds according to a two-state model (Egan *et al.*, 1993; Main *et al.*, 1999). The protein also auto catalyses refolding after denaturation with urea (Scholz *et al.*, 1996).

The unfolding temperature is characteristic of the protein and is defined by the mid-point of the melting process, T_m, when 50 % of the molecules are denatured. This can be thought of as the point when the free energy of the folded and un-folded forms are equivalent (Munoz and Sanchez-Ruiz, 2004). The end point of the melting process is more difficult to determine as unfolded protein has a tendency to aggregate and precipitate. T_m is dependent on buffer composition, pH and the presence of bound ligand (Cooper, 1999a). Unfolding of proteins is generally an endothermic process; the enthalpy required to break non-covalent interactions is the dominant term. Conversely, aggregation of unfolded molecules can be an exothermic process and can lead to noise in high temperature data.

DSC can be used as a screening technique for ligands. If a ligand preferentially binds to the folded or native form of a protein, the folded form of the protein is stabilised and T_m is increased, $+\Delta T_m$. If the ligand preferentially binds to the denatured form of the protein the protein is destabilised and T_m is reduced, $-\Delta T_m$ (Brandts and Lin, 1990; Matulis *et al.*, 2005; Cimmperman *et al.*, 2008). A disadvantage of DSC as a screening technique is the relatively high consumption of protein; in the case of FKBP12 700 µl of 80 µM (0.7 mg) is required. This was an important consideration with regard to both protein production and ligand solubility. Many ligands tested had maximum solubility in the order of 300 µM. The meant that the maximum protein ligand screening ratio was reduced to around 1:4 for this technique. This limited the possibility of detecting medium to low affinity ligands. Moreover, the instrument available did not have the capacity to test more than one sample at

a time or automate sample change over. Each sample took around 2 hours. DSC experiments were carried out to determine T_m for FKBP12, but not to screen ligands.

Materials and Methods

Differential scanning calorimetry (DSC) was used to determine T_m from changes in the specific heat capacity of the protein system with respect to temperature, relative to a buffer control. 79.7 μM FKBP12 in 50 mM ammonium acetate at pH 7 was heated from 30 to 80 ^oC with an increment of 0.1 ^oC in a Nano differential scanning calorimeter (TA Instruments – Waters LLC). A buffer baseline was determined by carrying out a similar experiment without protein.

Data Analysis

The DSC thermogram is a plot of partial molar heat capacity at constant pressure (C_p) against temperature, Figure 6-4. The DSC thermogram takes the form of a bell-shaped curve; T_m is determined from the maximum value of C_p. The enthalpy of unfolding, ΔH_u can be calculated from the area under the thermogram. ΔC_{pu}, the specific heat capacity change on unfolding, can be calculated from the difference in partial molar heat capacity before and after the melting transition. The magnitude of ΔC_{pu} is indicative of the relative increase in exposure of hydrophobic groups to solvent in the unfolded form of the protein. Data was processed using software supplied with the instrument and was normalised for protein concentration and corrected for buffer effects using a buffer baseline (C_pCalc, TA Instruments).

6.3.4 **Thermal denaturation fluorescence assay**

The thermal denaturation fluorescence assay (TDF) uses an environmentally sensitive fluorescent dye to follow temperature induced protein unfolding (denaturation). SYPRO orange (Invitrogen) was used to monitor the unfolding of a 50 µl of 10 µM FKBP12 (6 µg) in a 96-well format (Steinberg *et al.*, 1996). This represented a reduction in protein consumption of around 100 fold when compared to DSC. This method relies on the quantum yield of the dye being different in a hydrophobic or aqueous environment. Hydrophobic residues are largely buried within a folded protein; exposure of these residues on denaturation changes the quantum yield of the SYPRO dye in the environment of the exposed hydrophobic groups. Use of a real time PCR machine facilitated the excitation (λ_{ex} = 490 nm) of SYPRO dye and the detection of changes in fluorescent emission (λ_{em} = 590 nm) with temperature in a tightly controlled heat cycle (Pantoliano *et al.*, 2001; Epps *et al.*, 2001; Lo *et al.*, 2004). The TDF thermogram consists of a plot of fluorescent emission against absolute temperature. The midpoint melting temperature, T_m, can be estimated from the thermogram. As T_m is dependent on buffer stability, additives and pH, it was necessary to optimise conditions for screening. The aim was to choose a buffer system that had little pH dependence with temperature and control the use of additives that influenced T_m. As the background fluorescence of the SYPRO dye is influenced by the dielectric constant of the bulk solvent; it was important to control the % of solvent in all samples and use appropriate controls (Hawe *et al.*, 2008).

Materials and Methods

Standard screening conditions were 10 µM FKBP12 in 50 mM ammonium acetate, pH 7, 10x SYPRO on a 96-well PCR plate sealed with iCycler IQ™ optical tape (Bio-Rad). Ligands were screened at a concentration of 250 µM (unless otherwise stated). The total analyte volume was 50 µl. The SYPRO concentration is relative to the stock solution supplied by the manufacturer at 5000x in DMSO (equivalent to ~ 0.2 % DMSO final analyte). Experiments were carried out on a Bio-Rad iCycler IQ™ instrument with an iQ5 Real Time Detection System (V2.0 software, Bio-Rad). Protein melting was examined between 20 ^{o}C and 80 ^{o}C, using 0.5 ^{o}C temperature increment and 30 s hold at each temperature. In all screening experiments controls were included to measure T_m for an inhibitor free well (negative control) and for a known inhibitor SD1 (K_d = 7 µM, positive control) under identical buffering conditions. The negative and positive controls were always

performed in at least triplicate. The fluorescence of each compound was also investigated using a protein free control.

Estimating T_m by non-linear regression

The midpoint melting temperature, T_m, the enthalpy of unfolding, ΔH_u and the specific heat capacity of unfolding, ΔC_{pu} were calculated by fitting fluorescent emission thermal denaturation data to equation 22 by non-linear regression using the program Kaleidagraph (V4.03, Synergy Software). This equation describes the melting curve for the two-state reversible thermal denaturation of a protein (Pantoliano *et al.*, 2001). It is believed that the folding of FKBP12 is not unusual (Fulton *et al.*, 1999). It is one of the largest globular proteins thought to undergo reversible two-stage unfolding. FKBP12 was confirmed as a monomer in solution in this study (section 2.3.4). Floating a large number of unreliably estimated parameters can make non-linear regression algorithms non-convergent. There is also correlation between ΔH_u and ΔC_{pu} in the fitting and it is necessary to constrain one if a reliable estimate of the other is required. To minimise problems, an initial value for ΔH_u was entered from DSC data, ΔC_{pu} was selected for typical low molecular weight globular proteins (Cooper, 2005). It was necessary to truncate data soon after the protein had reached its maximum fluorescence, this is standard practice for this method of analysis (Lo *et al.*, 2004). The decrease in fluorescence is probably due to a number of reasons, including; protein aggregation and thermal quenching of SYPRO fluorescence (Ericsson *et al.*, 2006).

$$F(T) = F(\text{melt}) + \frac{[F(\text{fold}) - F(\text{melt})]}{1 + \exp\left\{\frac{-\Delta H_u^{T_0}}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right) + \frac{\Delta C_{pu}^{T_0}}{R}\left[\ln\left(\frac{T}{T_m}\right) + \frac{T_m}{T} - 1\right]\right\}}$$ **Equation 22**

Where: F(T) is the fluorescent emission at a given temperature, T; F(fold) is the baseline fluorescent emission before the melting-transition; F(melt) the fluorescent emission after the transition; $\Delta H_u^{T_0}$ the enthalpy of unfolding of the protein without the presence of the ligand (initial value for fitting was entered from DSC data, 6.3.3), $\Delta C_{pu}^{T_0}$ the heat capacity change on unfolding (starting parameter 2.2 $Kcal \cdot K^{-1} \cdot mol^{-1}$), T the temperature for a specific fluorescent emission, T_m the mid-point melting temperature and R the universal gas constant. All temperatures were absolute temperatures in Kelvin.

The experimental values for DSC and TDF are discussed in the TDF results section.

Estimating T_m from the thermal denaturation thermogram

The mid-point melting temperature, T_m, can also be estimated from the gradient of the denaturation thermogram. The maximum value of the first derivative of the fluorescent emission with respect to temperature was found to be a good approximation to T_m derived by fitting experimental data to equation 22, Figure 6-1.

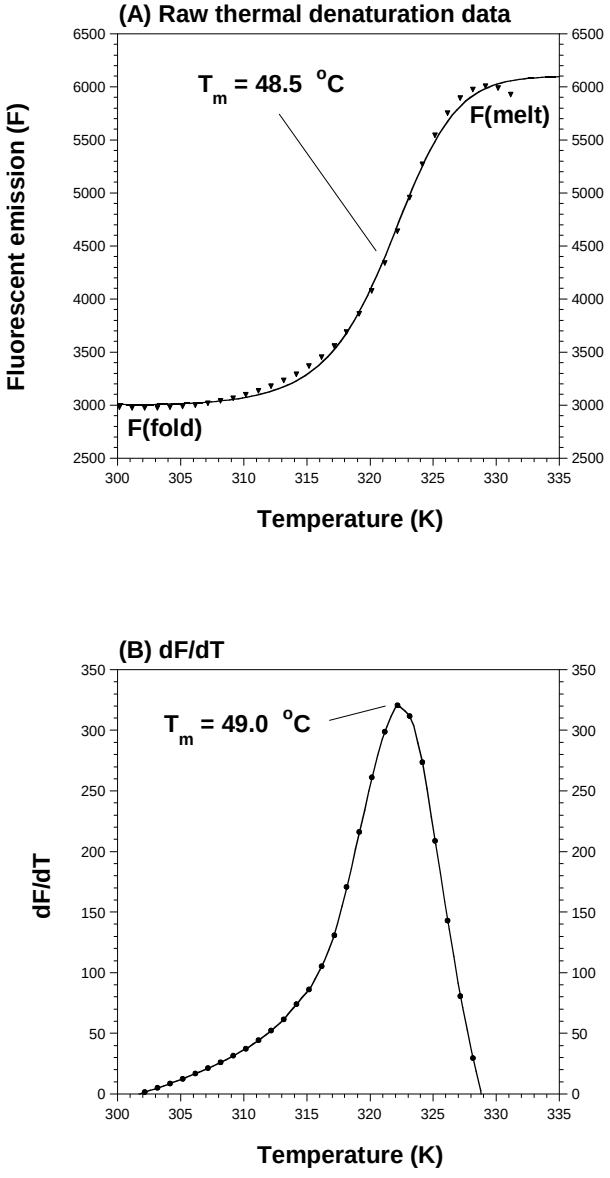


Figure 6-1 Estimation of T_m from a protein denaturation thermogram

(A) Typical raw fluorescent emission plotted as a function of absolute temperature. 10 μM FKBP12 in 10 mM sodium phosphate, pH 7; 10x SYPRO orange. Curve is the best fit of data to equation 22 using the non-linear regression program Kaleidagraph (V4.03, Synergy software). $\Delta H_u^{T_o}$ = 73.2±2.3 kcal·mol^{-1}, $\Delta C_{pu}^{T_o}$ = 3.69±0.15 kcal·K^{-1}·mol^{-1} (B) First derivative of the fluorescent emission in (A) with respect to temperature, dF/dT, plotted as a function of absolute temperature.

Estimating ligand affinity from shift in T_m

T_m was calculated with and without the presence of a ligand ($T_m = T_0$ with no ligand present) and the apparent equilibrium dissociation constant (K_d) of the ligand estimated at the melting temperature from equation 23 (Pantoliano *et al.*, 2001).

$$K_{dL\ T_m} = \frac{\left[L_{T_m}\right]}{\exp\left\{\frac{-\Delta H_u^{T_o}}{R}\left(\frac{1}{T_m}-\frac{1}{T_0}\right)+\frac{\Delta C_{pu}^{T_0}}{R}\left[\ln\left(\frac{T_m}{T_0}\right)+\frac{T_0}{T_m}-1\right]\right\}} \qquad \textbf{Equation 23}$$

Where: $K_{dL\ T_m}$ is the apparent equilibrium dissociation constant for the ligand at the melting temperature T_m; T_o the melting temperature of the apo protein; $\Delta H_u^{T_o}$ the enthalpy of unfolding for the apo protein; $\Delta C_{pu}^{T_0}$ the heat capacity change on unfolding; $\left[L_{T_m}\right]$ the concentration of the free ligand and R the universal gas constant.

Equation 23 makes the assumption that the concentration of the free ligand approximates to the total concentration of ligand, $\left[L_{T_m}\right]_{free} \rightarrow \left[L_{T_m}\right]_{total}$, when the ratio of ligand to protein is high.

It was possible to extrapolate the K_d of the ligand to an alternative temperature if thermodynamic data of the protein-ligand interaction was known from an ITC or SPR experiments. Knowledge of the enthalpy of binding of the ligand, ΔH_L, allows an estimation of the effect of temperature on K_d, equation 24. ΔH_L is only known if a detailed ITC or SPR study has been carried out. If this parameter was not available a bracketing approach was used (described at the end of this section).

$$K_{dL\ T} = K_{dL\ T_m}\ \exp\left\{\frac{-\Delta H_{L\ T}}{R}\left(\frac{1}{T}-\frac{1}{T_m}\right)\right\} \qquad \textbf{Equation 24}$$

Where: $K_{dL\ T}$ is the is the apparent equilibrium dissociation constant for the ligand at temperature T; $K_{dL\ T_m}$ is the apparent equilibrium dissociation constant for the ligand at the melting temperature T_m; T the temperature of interest; T_m the melting temperature with the presence of ligand and R the universal gas constant.

A better estimation is possible if the change in specific heat capacity on ligand binding, ΔC_{pL}, is known, equation 25.

$$K_{dL\ T} = K_{dL\ T_m} \exp\left\{\frac{-\Delta H_{L\ T}}{R}\left(\frac{1}{T}-\frac{1}{T_m}\right)+\frac{\Delta C_{PL}}{R}\times\left[\ln\left(\frac{T}{T_m}\right)+1-\frac{T}{T_m}\right]\right\} \quad \textbf{Equation 25}$$

Where: Variables are as defined for equation 22 and ΔC_{PL} is the specific heat capacity change on ligand binding.

The requirement of thermodynamic data to correct K_d for temperature is obviously a potential disadvantage for this method. A rank order of affinity can be creation at the protein melting temperature. The use of bracketing allowed the calculation of K_d at 25 ^oC for a range of putative ΔH_L. This approach was used in the screening experiments described in chapter 8. The K_d was extrapolated to 25 ^oC using 3 values for ΔH_L: -2000 cal·mol^{-1}, -5000 cal·mol^{-1}, -7000 cal·mol^{-1} and equation 24. ΔC_{PL} was not included in the analysis unless known (stated in the text). Another consideration is that a $+\Delta T$ can indicate binding at any site on the protein surface and not necessarily the site of interest.

The term apparent equilibrium dissociation constant, K_i, is particularly appropriate in the TDF studies. The magnitude of the equilibrium dissociation constant estimated by this method is conditional upon the % of DMSO and any pH change caused by addition of ligand. All experiments were carried out with a minimum concentration of 0.2 % DMSO.

Screening the test set by TDF assay

The TDF assay was used to screen members of the inhibitor test set to assess the threshold affinity of the assay. A ligand screening compound concentration of 300 μM was chosen and the experiment repeated 6 times (method was as described above). The K_d was calculated at the screening concentration using equation 24 and the bracketing technique described above.

6.3.5 **Surface plasmon resonance**

Surface plasmon resonance (SPR) can be used to investigate the affinity of a ligand for a protein and provide kinetic information (Cooper, 2002; Navratilova *et al.*, 2007). The SPR phenomenon allows for the indirect measurement of the proportion of protein in complex with a ligand in real-time. The sensor "chip" has two faces; a thin gold film backed by glass. Light is incident on the inner surface of the gold through the glass, analyte solutions are passed over the exposed gold surface. The exposed surface of the gold film is coated to enable protein to be immobilised by means of a coupling reaction. The choice of coupling chemistry is dependent on the protein under investigation. It is important that the coupling reaction does not lead to occlusion of the protein binding site and that the macromolecule retains its native affinity.

Ligand/reference solutions are passed over the immobilised protein by means of a flow cell. When a light is incident on the gold film there is a resonance energy transfer between the incident light and free electrons in the metal. Energy transfer is dependent on the refractive index of the media in close proximity to the surface. Refractive index, a function of density, is dependent on ligand occupancy of the immobilised protein. If the reflected intensity is measured as a function of the reflected angle, the SPR effect manifests as a dip in the intensity of the reflected light at a specific angle, the resonant angle. Energy transfer is measured by changes in the reflected light and is output as response units (RU). Ligand passing over the immobilised protein binds, and attains equilibrium with the protein at a measurable rate. This rate can be correlated to the rate of association of the ligand with the protein. The rate of dissociation of the ligand can be determined by monitoring rate of reduction in the response units as a ligand free solution is passed over the complex. It is possible to calculate the affinity of the ligand from the magnitude of the steady-state response over a ligand concentration series. This was the method used in this study. It is usual to compare the SPR of the immobilised protein of interest with a surface prepared in an identical fashion minus the protein component, the reference surface. Identical fluids are passed over each surface during the experiment.

FKBP12 was found to be inactive following coupling via primary amines (CM5 chip). Coupling via lysine amines was thought to lead to occlusion of the binding site. FKBP12 has been successfully coupled by Ni^{2+}-nitrilotriacetic acid (NTA)/His-tag capture (NTA chip) and streptavidin-biotin-X-NTA (streptavidin chip, X is a 10-carbon spacer) in this laboratory

(Wear and Walkinshaw, 2007). The study by Wear describes how kinetics and affinity data were acquired for the FKBP12-rapamycin complex. Thermodynamic parameters are in agreement with ITC studies.

Materials and Methods

Data was collected for FKBP12 immobilised on an NTA chip on a Biacore T100 instrument (GE Healthcare). Protein was captured on the Ni^{2+}-NTA surface followed by brief covalent stabilisation using the method described by Wear *et al.* (2007). This method has been shown to produce surfaces of between 65 and 90 % activity with minimal baseline drift. Ligand was passed over the chip at concentrations in the range 0 to 50 μM ligand (SD1, SD12, SD2 and SD3); 137 mM sodium chloride; 2.7 mM potassium chloridel; 0.005 % P20 surfactant; 10 mM sodium phosphate, pH 7.3 at a flow rate of 30 $\mu l \cdot min^{-1}$. The experiment was carried out at 25 ^{o}C for all ligands and repeated at between 10 and 40 ^{o}C (5 ^{o}C intervals) for SD1. Experiments were repeated in triplicate. Data was corrected for bulk buffer effects by subtracting the response from the reference cell and the baseline was set to zero.

Estimating affinity from steady state affinity data

The steady-state response (R_{eq}) was plotted against the concentration of ligand. Data was fitted to a Langmuir 1:1 binding model, equation 26, using the software supplied with the instrument (Biacore T100 software V1).

$$R_{eq} = \frac{R_{max}}{(1+\left(\frac{K_d}{L}\right)} \qquad \textbf{Equation 26}$$

Where: R_{eq} the response units after baseline correction at steady-state for a specific ligand concentration, R_{max} the maximum response at protein saturation, K_d the equilibrium dissociation constant and [L] the concentration of the ligand.

Van't Hoff analysis

The van't Hoff method was used to measure ΔH from the dependence of K_d on temperature, equation 27. $\ln K_d$ was plotted against 1/T.

$$\ln K_d = \frac{\Delta H}{RT} - \frac{\Delta S}{R} \qquad \textbf{Equation 27}$$

Where: K_d is the equilibrium dissociation constant, ΔH the enthalpy change on binding, ΔS the entropic change on binding, T the absolute temperature and R the universal gas constant.

For data exhibiting a linear van't Hoff relationship; the gradient of the graph is equal to $\frac{\Delta H}{R}$ and the intercept with the y-axis, $\frac{\Delta S}{R}$. Curvature in the graph suggests that ΔH is temperature dependent (Naghibi *et al.*, 1995). Including an additional term in the equation accounts for the specific heat capacity change on ligand binding at constant pressure, ΔC_p, this is described by equation 28 (an integrated form of the van't Hoff equation considering a reference temperature).

Non- linear van't Hoff plot – including ΔC_p

Equation 28 was fitted to non-linear van't Hoff plots using least squares non-linear regression (Kaleidagraph V4. 03, Synergy Software). It is often difficult to detect non-linearity in a van't Hoff plot, due to uncertainty in the data, particularly at high temperatures, and the narrow temperature range explored in the experiments (Chaires, 1997). Equations 27 and equations 28 were both fitted to data to determine the most appropriate methodology for the protein-ligand system under investigation.

$$\ln K_d = \frac{1}{R}\left\{\frac{\Delta H_{ref}}{T} - \Delta S_{ref} + \Delta C_p\left[\frac{T - T_{ref}}{T} - \ln\left(\frac{T}{T_{ref}}\right)\right]\right\} \qquad \textbf{Equation 28}$$

Where: variables and constants are those described for equation 27 at the reference temperature; ΔC_p is the specific heat capacity change on ligand binding at constant pressure and T_{ref} = 298.15 K ≡ 25 ^oC.

6.4 Screening the ligand test set - results and discussion

6.4.1 Peptidyl-prolyl isomerase assay

Determining threshold affinity using the ligand test set

Estimating the affinity of a ligand for FKBP12 using the PPIase method and fitting data to inhibition curve described by equation 20 was extremely low throughput; taking ~ 1 week per ligand. This was useful as a final characterisation step for a specific FKBP12-ligand complex, but too slow for use as a screening strategy for virtual hits.

A single point assay was tested as a method to rank a ligand series by affinity. It was well known that there is considerable variability in PPIase results for identical solutions for the reasons described in the methods section. A threshold affinity was determined for the PPIase assay for the ligand test set at a compound screening concentration of 300 μM repeated 9 times. This approach increased throughput to 3 compounds/day. Figure 6-2 shows the results for a single point assay of 300 μM ligand (SD1, SD2, SD12 and SD3) at 4 ^{o}C . FK506 (K_d = 0.4 nM) was used as the positive control. Results are presented as the relative rate of turnover of substrate relative to the turnover of the enzyme under identical conditions with no inhibitor present. The positive control, FK506, gave 98 % inhibition at a concentration of 0.5 μM. FK506 is not particularly soluble in aqueous buffer and could not be tested at the same concentration as the test set compounds. Inhibition was seen for compounds with $K_d \leq$ 157 μM (SD3, 6 % inhibition), lower affinity ligands were not seen to inhibit the turnover of the enzyme at this screening ratio. Even with nine repeats the standard deviation in the % inhibition was ±10 %. Compound SD12 appeared above SD2 in the rank order by this method. However, the difference in mean rate of substrate turnover between the two ligands was in the same order as the standard deviation of 9 repeats. The ligands did not significantly affect the background turnover of the enzyme.

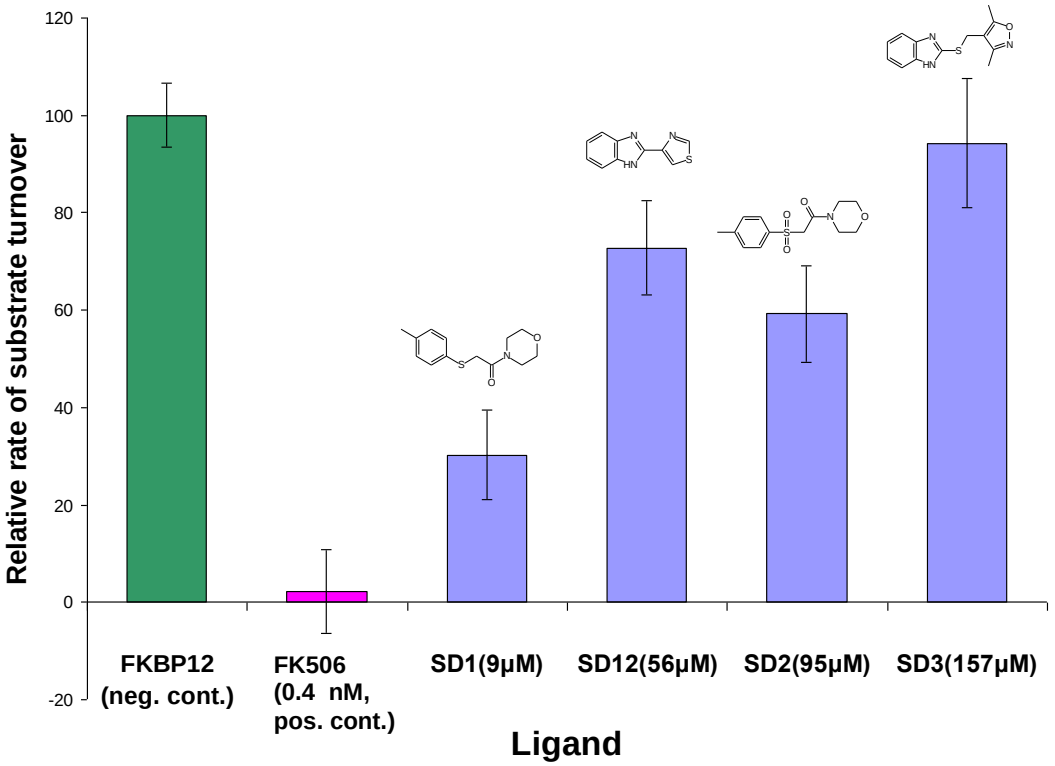


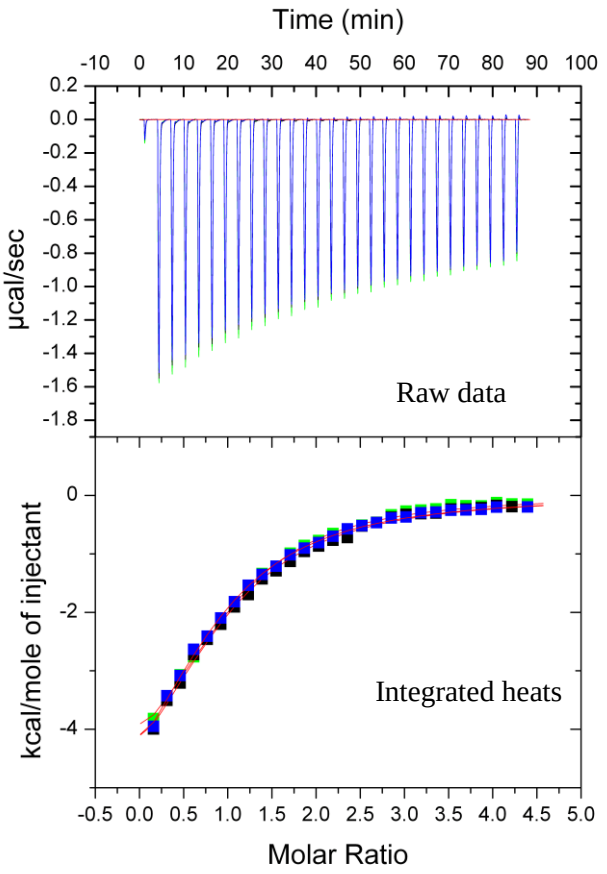
Figure 6-2 The ligand test set and PPIase inhibition

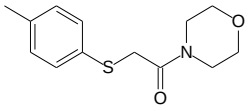
Screening the ligand test set at a single concentration, error bars show standard errors (n=9). Rates were corrected for background turnover of the substrate at 4 ^{o}C. 100 nM FKBP12, 300 µM ligand (SD1, SD2, SD12 and SD3), 0.6 $mg^{-1} \cdot ml^{-1}$ α-chymotrypsin, 120 µM Succ-Ala-Leu-Pro-Phe-*p*-nitroaniline, 14 mM lithium chloride, 1 % methanol (v/v), 3 % 2,2,2-trifluoroethanol, PBS, pH 8, 4 ^{o}C. FK506 was used as the positive control at 0.5 µM. Structures are shown for interest for the SD series of ligands.

It is worth noting that this method is not tolerant of precipitation in the analyte. Even low levels of precipitate reduce the A_{400nm} and can lead to false positives.

6.4.2 Isothermal titration calorimetry

Figure 6-3 shows representative ITC data for the FKBP12-SD1 complex. 300 µM SD1 was injected into 13.7 µM FKBP12 at 25 ^{o}C; 50 mM sodium chloride; 50 mM sodium phosphate, pH 6.8. Heats of ligand dilution were relatively large for SD1. This was apparent in the raw data and was accounted for in the data analysis. SD1 was the lowest affinity ligand in the test set that showed measurable affinity at these screening conditions.





SD1

K_d = 6.8 ± 1.2 µM

ΔG = -7.1 ± 0.1 kcal·mol^{-1}

ΔH = -6.1± 0.6 kcal·mol^{-1}

TΔS = 1.0± 0.7 kcal·mol^{-1}

N = 0.96 ± 0.05

Figure 6-3 Determining K_d of the FKBP12-SD1 complex by SD1

(Top panel) Raw ITC data. 0. 3 mM SD1 was injected into 1.47 ml, 13.7 µM FKBP12 in 28 x 10 µl injections (0.5 µl 1st injection), 180 s apart at 25 ^{o}C. All reagents were dissolved in 50 mM sodium chloride; 50 mM sodium phosphate, pH 6.8.
(Bottom panel) Integrated heat data was corrected for heat of ligand dilution and fit to a standard 1:1 binding model, red curves (MicroCal Origin software, V4.0). The inset show the structure of SD1, the thermodynamic profile and the stoichiometry of the interaction, N. Results are in triplicate (black, green blue), standard errors are shown between repeats (over 3 days) rather than the error in the fit to the data.

It is interesting to note that although the data shown for the 3 repeats appears to be very reproducible, fits the model well and gives a stoichiometry (N) close to 1, there was still a reasonable degree of uncertainty in the thermodynamic parameters. The uncertainty is calculated for repeats on three different days rather that the fit of the model to the data. These results are typical for ITC where ΔG is estimated with the greatest precision, followed by ΔH. TΔS has the highest uncertainty as it is calculated arithmetically from ΔG and ΔH (Holdgate, 2001). For SD1, a relatively low affinity ligand, the curve of integrated heats plotted as a function of protein/ligand molar ratio is not sigmoid. Increasing the protein

concentration would address this issue to some degree and "steepen" the curve, although this would make the assay more protein hungry. Raising the concentration of the ligand is not possible because of solubility issues. These results indicate that the interaction between FKBP12 and SD1 is both enthalpically and entropically driven (this is discussed further in Chapter 4).

Results for members of the test set are shown in the combined results Table 6-6. The lowest K_d detected with these screening conditions was 6.8 μM (SD1). Compounds SD12, SD2 and SD3 all showed measurable heats of dilution. Enthalpies of binding were too small and too linear to process at these analyte concentrations. The closest affinity ligand in the test set to SD1 is SD12 with a predicted K_d = 56 μM. ESI-MS suggests that this compound unfolds FKBP12 rather than forming a non-covalent interaction. SD2, K_d = 95 μM is of much lower affinity than SD1.

The activity of the protein was tested after 1.5 weeks storage in the ITC hotel (4 ^{o}C) and was found to be > 95 % of the starting activity by PPIase assay. FKBP12 was therefore considered to be a suitable protein for auto-ITC experiments over a long timescale.

6.4.3 Differential scanning calorimetry

The thermal denaturation of FKBP12

FKBP12 is a single domain protein and the DSC thermogram showed a simple bell-shaped curve. T_m for FKBP12 was estimated to be 53 ^{o}C; Figure 6-4. The enthalpy of unfolding, ΔH_u of FKBP12 was determined from the area under the DSC thermogram curve. Data was processed using software supplied with the instrument and was normalised for protein concentration and corrected for buffer effects using a buffer baseline (CpCalc, TA Instruments). Ammonium acetate showed good thermal stability at the temperatures employed. The pKa of ammonium acetate changes from 4.7 to 4.8 between 20 and 80 ^{o}C respectively (TA instruments data table). The unfolding of FKBP12 was an endothermic process with an enthalpy of unfolding, ΔH_u = 42.7 kcal·mol^{-1} and an associated entropic change, ΔS_u = 131 cal·K^{-1}·mol^{-1}. The value for ΔH_u is considerably lower than that obtained in the analysis of TDF data (ΔH_u = 73.2 ± 2.3 kcal·mol^{-1}). Cooper gives a typical value for a small globular protein of 71.8 kJ·mol^{-1} (300 kJ·mol^{-1}) (Cooper, 1999b). The entropic change is small compared to Cooper's typical value of ΔS_u = 229 cal·K^{-1}·mol^{-1} (3.2 kJ·K^{-1}·mol^{-1}). The change in specific heat capacity, ΔC_p = 2.7 ± 0.15 Kcal·K^{-1}·mol^{-1} calculated in the TDF

analysis, is larger than Cooper's typical value of 2.2 $Kcal \cdot K^{-1} \cdot mol^{-1}$ (9 $kJ \cdot K^{-1} \cdot mol^{-1}$). The DSC was only available for a short time. It would be interesting to repeat these experiments.

Addition of 2 % DMSO (286 mM) increased the melting temperature to 53.8 ^{o}C, $\Delta T_m = 0.8\ ^{o}C$. This increased thermal stability of the protein was consistent with the findings that DMSO binds to the active site of FKBP12 with an affinity in the tens of millimolar range (Burkhard *et al.*, 2000).

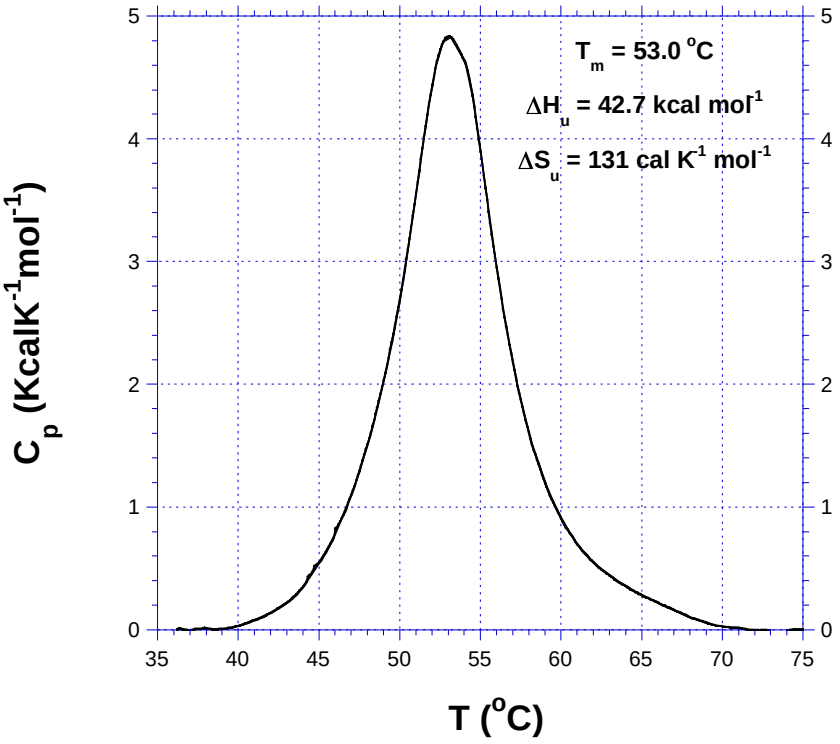


Figure 6-4 DSC thermogram of FKBP12

DSC thermogram for 79.7 μM FKBP12 in 50 mM ammonium acetate, pH 7. The protein had a mid-point melting temperature of 53 °C. Nano DSC (TA Instruments – Waters LLC). Data was normalised for protein concentration and corrected for buffer effects by using the buffer baseline. T_m and the enthalpy and entropy of unfolding included on the figure. Dr Daphne Kan from TA Instruments assisted in processing the data.

6.4.4 Thermal denaturation fluorescence assay

Optimising screening conditions for FKP12

Figure 6-5 shows the influence of pH, protein and solvent concentration on T_m. Figure 6-5 (A) shows that pH had a significant effect on T_m. All screening experiments were carried out at pH 7. These results are in agreement with experimental data from ESI-MS experiments where low pH was found to give a spectrum consistent with a denatured protein (section 2.3.3). A reduction in T_m with more acidic conditions is typical for small globular proteins (Cooper *et al.*, 2001).

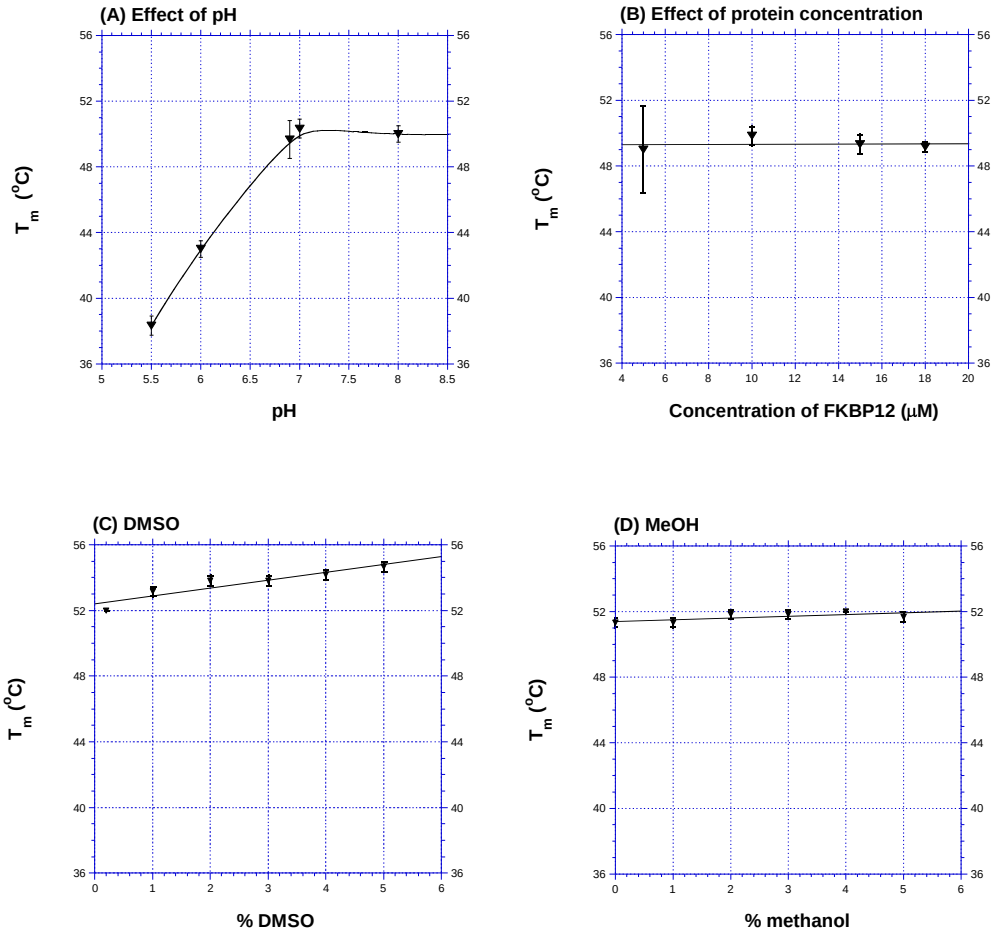


Figure 6-5 Optimising screening conditions

(A) T_m showed a correlation with pH. 18 μM FKBP12. (B) T_m does not show a dependence on protein concentration. However, the variance in T_m was correlated to protein concentration, PBS, pH 7. (C) Increasing the % of DMSO increased T_m. 18 μM FKBP12, 0 - 5 % DMSO, PBS, pH 7. (D) Low concentrations of methanol had no significant effect on T_m. 18 μM FKBP12, 0 - 5 % methanol, PBS, pH 7. Error bars are standard deviations for n = 3. "Straight line" curves are best fit to the data (y = mx + c); the curve shown on part (A) is to aid visualisation.

The estimation of $T_{m.}$ was not dependent on protein concentration. The lack of correlation of T_m with concentration suggests that FKBP12 is not influencing folding/unfolding in its capacity as a peptidyl-prolyl isomerase, in this concentration range. Figure 6-5 (B) shows how the variance of 3 repeats reduced as the protein concentration was increased above 5 μM. 10 μM FKBP12 was used in all screening experiments; this gave a large enough change in fluorescence between folded and unfolded states. Fitting data to thermograms gave smaller, more random residuals when the difference between the F(fold), the baseline fluorescent emission before the melting-transition, and the F(melt), the fluorescent emission after the transition, was greater than 1500 fluorescence units. Observations from within our group suggest there is inverse relationship between the molecular weight of the protein investigated and minimum molarity required for the protein in the analyte solution.

Solvents were kept to a minimum by preparing solvent stock solutions of ligand at high concentration. DMSO was found to stabilise the protein at low concentrations, Figure 6-5 (C). This is consistent with previous studies; where DMSO was found to inhibit the active site of FKBP12 with a K_d of ~20 mM (Burkhard *et al.*, 2000). Jackson *et al.* found that trifluoroethanol (3.6 - 9.6 % (v/v)) stabilised the native state of FKBP12 to urea-induced denaturation (Main *et al.*, 1999). Methanol did not significantly influence T_m between 0 and 5 %. Compounds were routinely dissolved in methanol, unless obtained as stock solutions in DMSO. DMSO was incompatible with ESI-MS and was also avoided for this reason.

SD1 as a positive control for FKBP12

Compound SD1, K_d = 6.8 μM, was used as a positive control in thermal denaturation experiments at a concentration of 250 μM. SD1 binds to FKBP12 with 1:1 stoichiometry and competes with a preferred substrate for the active site of the enzyme. The ligand is soluble to 100 mM in methanol.

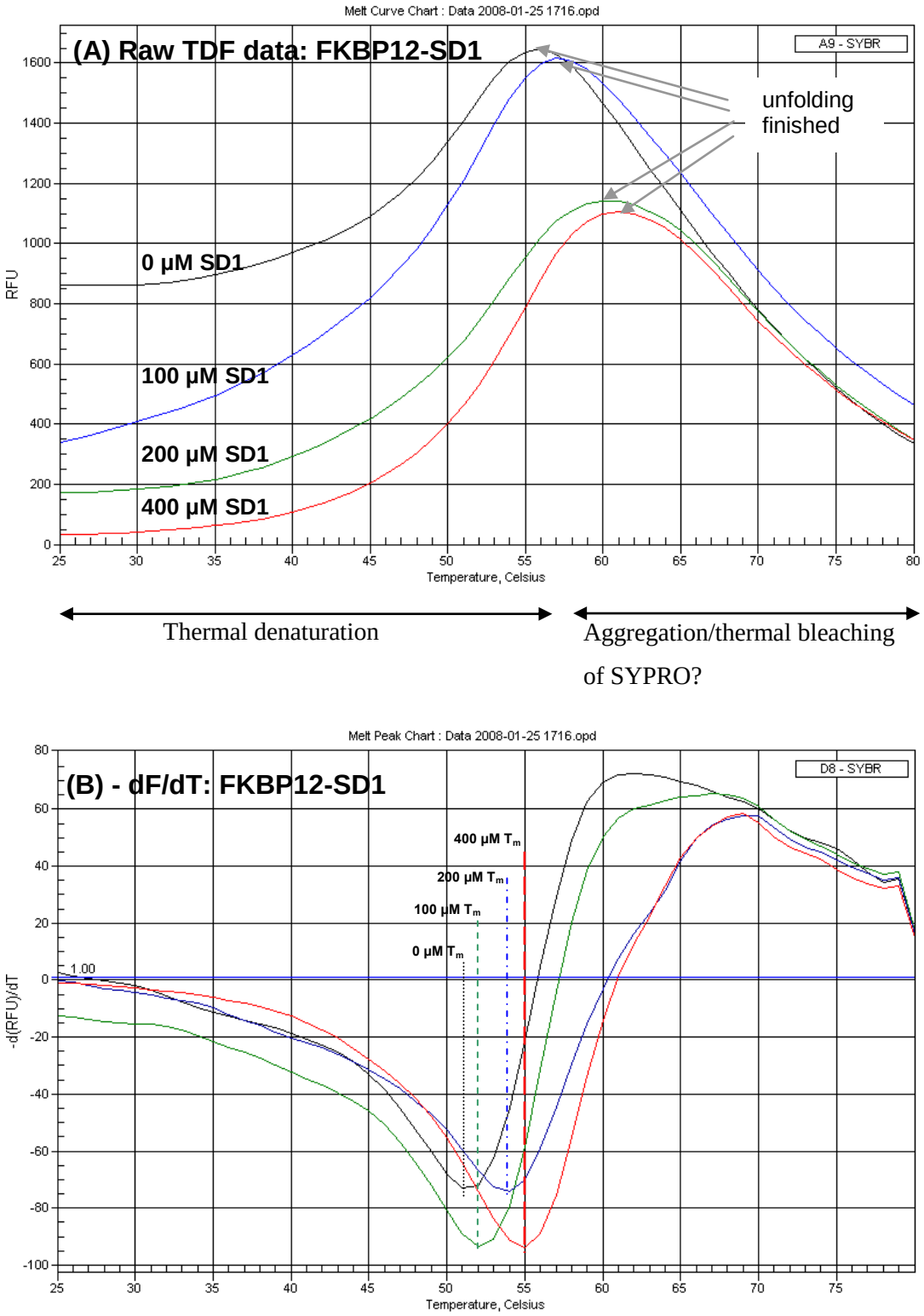


Figure 6-6 TDF FKBP12-SD1

(A) Typical raw TDF data. Melting curves are shown for 10 µM FKBP12 at 0 (black), 100 (green), 200 (blue) and 400 µM (red) SD1 respectively, 50 mM ammonium acetate, pH 7; 10x SYPRO, 1% methanol. Thermal denaturation occurs between 30 and 60 ^{o}C. The decrease in fluorescence after melting is thought to be due to aggregation/thermal bleaching. (B) –dF/dT for graph (A). The melting temperature T_m increases with increasing concentration of ligand. T_m is labelled for SD1 between 0 and 400 µM.

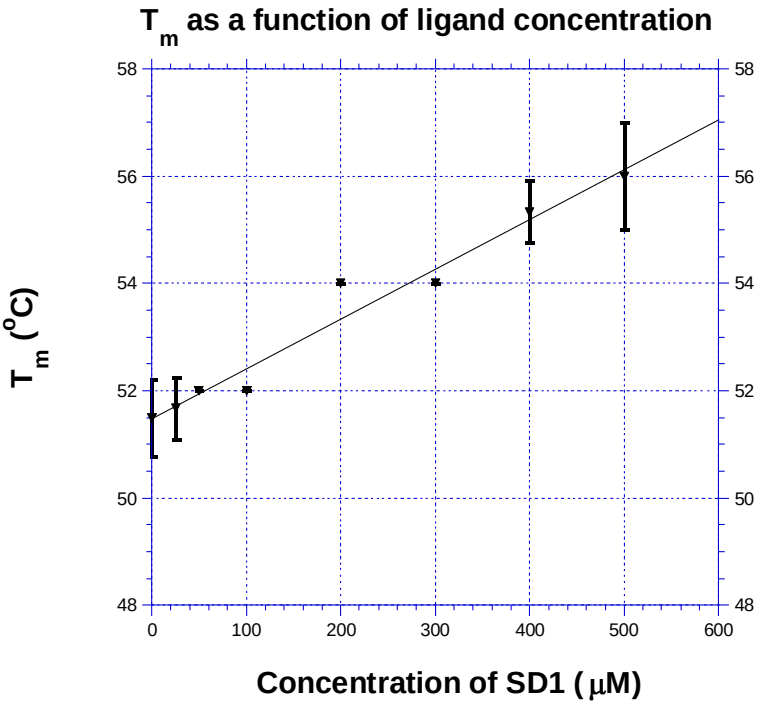


Figure 6-7 SD1 as a positive control

T_m as a function of the concentration of SD1; error bars show standard deviations for results in triplicate. 10 μM FKBP12 at 0, 100, 200 and 400 μM SD1 respectively, 50 mM ammonium acetate, pH 7.0; 10x SYPRO, 1% methanol.

K_d estimated by thermal denaturation fluorescence were similar to values obtained from ITC (section 6.4.2), Table 6-3. The value quoted has been extrapolated to 25 ^oC using equation 25 as thermodynamic parameters are known for the FKBP12-SD1 complex (ITC 6.4.2, SPR 6.4.5). Data for DMSO is included in the table for interest. For this compound the data has been extrapolated to 25 ^oC using thermodynamic data estimated from buried surface area calculated from x-ray crystallographic data (Burkhard *et al.*, 2000). Buried surface area = 41.7 $Å^2$. (1d7h.pdb, 1.9 Å resolution). $\Delta H_{L(ITC)}$ = - 0.17 $cal^{-1}{\cdot}mol^{-1}{\cdot}Å^{-2}$ (Wear *et al.*, 2007a) and ΔC_p = - 0.4 $cal^{-1}{\cdot}K^{-1}{\cdot}mol^{-1}{\cdot}Å^{-2}$ (Connelly and Thomson, 1992). For these reasons the result for DMSO must be treated with some caution. However, it is in general agreement with fluorescence data reported by Burkhard.

Ligand	*K_d (μM)*		
	TDF ***25 °C***	***ITC*** ***25 °C***	***SPR*** ***25 °C***
SD1	7.2±2.6	6.8±1.2	29.8±0.8
DMSO	37000±14000	-	49000±19000

Table 6-3 Comparison of K_d estimated by TDF with other biophysical techniques

Table shows K_d calculated from TDF, ITC and SPR assays. Standard deviations are shown for at least n=3 (Affinity measured by TDF has been transformed to 25 ^oC using equation 25 (For SD1 $\Delta H_{L(ITC)}$ = -6.09 kcal·mol^{-1}, $\Delta C_{p(spr)}$ = -712 cal·K^{-1}·mol^{-1}. For DMSO $\Delta H_{L(calc)}$ = -7.09 kcal·mol^{-1}, $\Delta C_{p(calc)}$ = -270 cal·K^{-1}·mol^{-1}).

Influence of inhibitor on pre-transition fluorescence, F(fold)

FKBP12 has large shallow hydrophobic pocket, containing many aromatic residues. This forms the binding site for the immunosuppressants FK506 and rapamycin, and proline in peptide substrates. It is interesting to note that F(fold) is reduced on addition of an inhibitor in a dose dependent manner, to a value close to zero at saturation (Figure 6-7).. The data can be rationalised by SYPRO orange having an increased quantum yield in the environment of the active site. The presence of an inhibitor in the active site, with a higher affinity to FKBP12 than that of SYPRO, displaced SYPRO into the bulk solvent and fluorescence decreased. SYPRO orange has a commercially sensitive structure and MW; it was not possible to quantify binding. Similar effects were obtained in a series of experiments using the environmentally sensitive dye 1-anilinonaphthalene-8-sulfonate (ANS), (Cardamone and Puri, 1992). It was not possible to reach saturation when ANS was titrated with FKBP12 (results not presented).

Determining threshold affinity using the ligand test set

Figure 6-8 shows ΔT_m for a screen of the ligand test set: 300 μM ligand (SD1, SD2, SD12 and SD3), 2 μM ligand (rapamycin and FK506), pH 7, repeated 6 times. The bar graphs were divided into (A) and (B) to distinguish the nanomolar affinity ligand (rapamycin and FK506) protein/ligand screening ratio (0.2) from the micromolar affinity ligand (SD1, SD12, SD2 and SD3) screening ratio (25). Experiments were carried out with the same batch of protein on the same 96-well plate. The same negative control is shown on both bar charts labelled protein mean. The ligand test set was particularly useful for this assay as the compounds are not fluorescent at the excitation/emission wavelengths of SYPRO orange.

FK506 and rapamycin are only soluble to 2 μM in aqueous buffer. At this protein/ligand ratio (0.2), ΔT_m was small compared to the standard error (n = 6) and was felt to be too small to use these compounds as a positive control. Rapamycin appears to stabilise FKBP12 more than FK506, although the difference is small compared to the standard error and cannot be considered significant. The estimated K_d for rapamycin were 80 ± 40 nM and 130 ± 40 nM respectively when extrapolated to 25 ^oC ,Table 6-4. These estimates are over a factor of 10 higher than published K_d. This in part due to the estimation of K_d from the model described in equation 22. This model is only a good approximation if $\left[L_{T_m}\right]_{free} \rightarrow \left[L_{T_m}\right]_{total}$; this is not the case for FK506 and rapamycin.

In the micromolar affinity group, shown in Figure 6-8 (B), the assay discriminated between SD1 (K_d = 7 μM), the only low micromolar affinity ligand and the other members of the group; SD12 (K_d = 56 μM), SD2 (K_d = 95 μM) and SD3 (K_d = 157 μM). SD1 gave a ΔT_m = 4.7 ± 0.6 ^oC at 300 μM. SD3 showed the second largest ΔT_m in this set, leading to an over estimation of K_d, Table 6-4. SD3 was not confirmed as an inhibitor by ITC and appeared as only a weak inhibitor in the PPIase assay. Interestingly SD12 produced a positive ΔT_m, this is usually interpreted as a compound preferentially binding to the folded form of the protein (Cimmperman *et al.*, 2008). The ligand can of course bind to different conformations present in the assembly with varying affinity. The melting curve only shows the global change in fluorescence. ESI-MS showed that SD12 appeared to destabilise FKBP12 in a time-dependent manner and give a spectrum consistent with an unfolded protein. There was no evidence in ESI-MS that SD12 bound to the unfolded or folded form.

SD1 was chosen as the positive control for the TDF assay due to its good solubility and the large ΔT_m at the screening concentration (250 μM). Extrapolating to 25 ^oC requires an estimation of the enthalpy of binding of the compound. The enthalpy of SD1, ΔH = -6 kcal mol^{-1}, this was thought to be more representative of the small molecule inhibitors from the screening assays than the large immunosuppressive ligands that have $\Delta H < -12$ kcal·mol^{-1}.

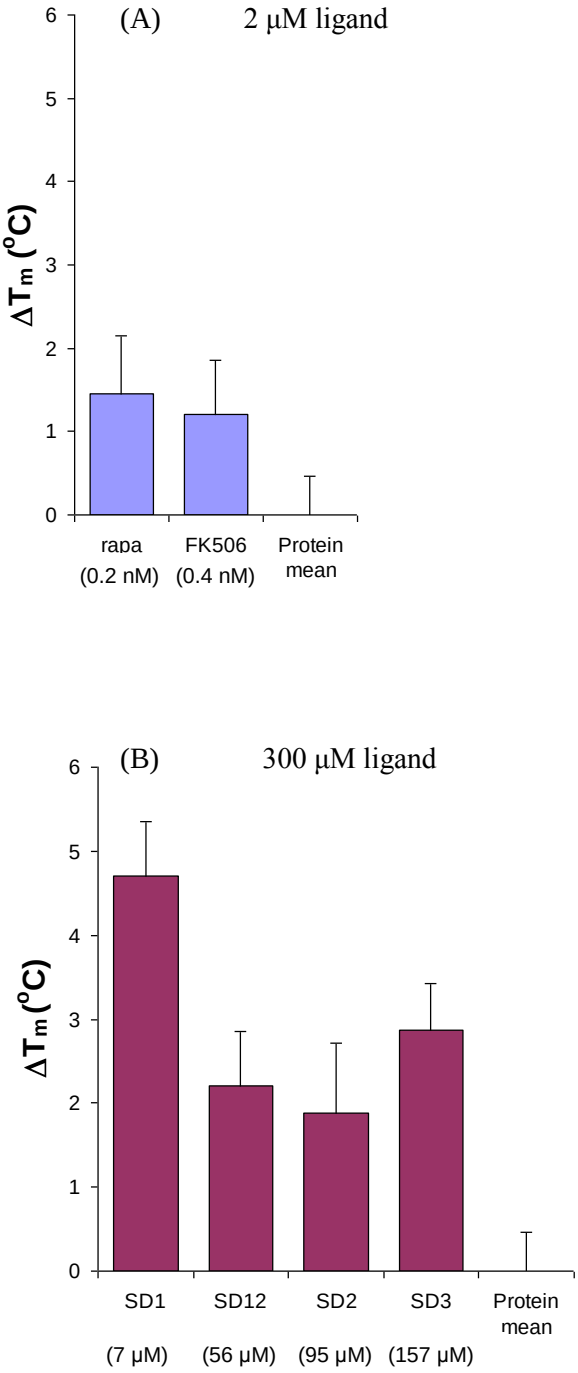


Figure 6-8 TDF for the ligand test set

The bar graphs show the shift in melting temperature for the ligand test set at a single concentration. 10 µM FKBP12, 50 mM ammonium acetate, pH 7; 1 % methanol (v/v). (A) 2 µM (rapamycin, FK506), (B) 300 µM (SD1, SD2, SD12 and SD3), error bars show standard errors (n = 6).

Ligand	K_d µM T_m	K_d µM 25 °C	K_d µM ΔH^a -2.0 kcal·mol^{-1} 25 °C	K_d µM ΔH^b -5.0 kcal·mol^{-1} 25 °C	K_d µM ΔH^c -7.0 kcal·mol^{-1} 25 °C	K_d µM	ΔH_L kcal·mol^{-1}	ΔC_{pL} kcal·K^{-1}·mol^{-1}	Reference (therm. param.)
rapa.	**1.2±0.6**	0.08±0.05	*0.9*	*0.6*	*0.5*	**0.0005$_{ITC}$**	-16.4	-0.30	(Connelly and Thomson, 1992)
FK506	**1.3±0.6**	0.13±0.04	*1.0*	*0.7*	*0.5*	**0.001$_{ITC}$**	-14.2	-0.26	(Connelly and Thomson, 1992)
SD1	**50±28**	4.42±3.48	*37.2*	*23.7*	*17.5*	**6.8$_{ITC}$**	-6.09	-0.71	EAB, section 6.4.2
SD12	**133±68**	-	*101.3*	*66.8*	*50.7*	**56$_{nmr}$**	-	-	(Stebbins *et al.*, 2007)
SD2	**151±93**	-	*115.2*	*76.4*	*58.1*	**95$_{nmr}$**	-	-	(Stebbins *et al.*, 2007)
SD3	**103±49**	-	*78.0*	*51.0*	*38.4*	**157$_{nmr}$**	-	-	(Stebbins *et al.*, 2007)

Table 6-4 Estimating $K_{d(app)}$ from TDF results

The K_d for the ligand complex was estimated at T_m (equation 23); $\Delta H_u^{T_o}$ = 73.2 kcal·mol^{-1}, $\Delta C_{pu}^{T_o}$ = 2.75 kcal·K^{-1}·mol^{-1}. If thermodynamic information was available K_d was extrapolated to 25 °C (equation 25). For screening, when thermodynamic parameters were not available, K_d were calculated for three putative enthalpy, ΔH^{a-c}: -2 kcal·mol^{-1}, -5 kcal·mol^{-1}and -7 kcal·mol^{-1} (results in italics). Standard errors are for values on separate days/plates.

Table 6-4 illustrates how the extrapolation of K_d at T_m to 25 °C is influenced by ΔH_L and ΔC_{pL}. For compound SD1, where thermodynamic parameters had been recently measured and the constraints of the biophysical model are met, TDF gives a good estimation of affinity. This was also true for compound KB2_61D (7 µM screening hit) and is discussed in Chapter 8. Use of K_d at T_m (~ 55 °C) and a single point assay is probably sufficient to distinguish nanomolar (rapamycin, FK506), low micromolar (SD1, KB2_61D) and medium micromolar affinity ligands (SD12, SD2, SD3). For the nanomolar ligands the three putative binding enthalpies are much smaller than those measured by ITC and extrapolations using these values are poor estimates of K_d measured by other methods at 25 °C.

The assay in this form provides a medium through-put screen. It would be possible to screen 96 compounds and controls (positive/negative and controls for screening compound fluorescence) in 12 hours.

Uncertainty in estimating K_d by TDF

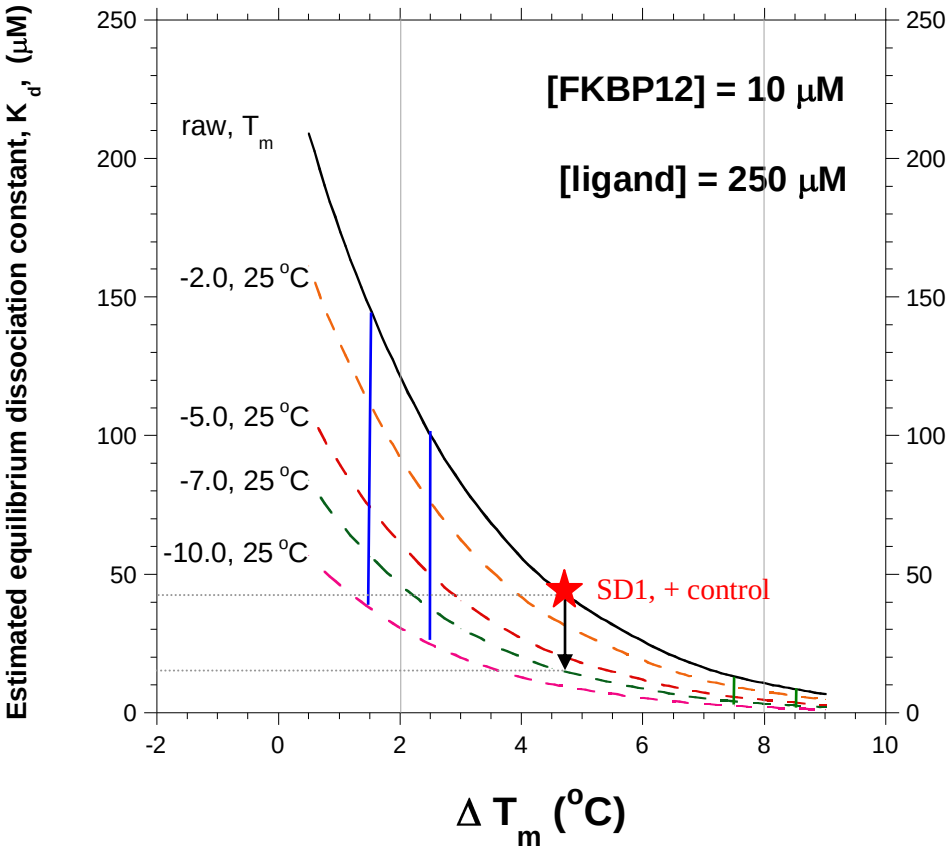


Figure 6-9 Uncertainty in estimated K_d is correlated to both estimated binding enthalpy and the magnitude of ΔT_m

The effect of using an incorrect binding enthalpy is more pronounced for lower affinity ligands, showing small ΔT_m. The figure shows the effect of extrapolating raw TDF data (black curve) for a series of putative enthalpies (dashed curves: -2 to -10 kcal·mol^{-1}) to 25 °C. The extrapolation was simulated using the model described by equation 24; as would be the case for a screening experiment. This model does not consider a temperature dependence on enthalpy. Considering ΔC_p would further increase the uncertainty in the estimation for K_d. The blue and green bars define areas of uncertainty for ΔT_m = 2 ± 0.5 °C and 8 ± 0.5 °C respectively for enthalpy variable between -2 and -10 kcal·mol^{-1}. The red star represents the extrapolation of the raw estimate of K_d from the mid-point melting temperature to 25 °C for SD1, the positive control for the TDF experiments. The simulation considers [FKBP12] = 10 μM and [ligand] = 250 μM.

The effect of using an incorrect binding enthalpy in the extrapolation to 25 °C must be viewed in the context of the experimental error in T_m (~ 0.5 °C) and the magnitude of ΔT_m, Figure 6-9. Low affinity ligands have a small value of ΔT_m compared to high affinity ligands for a given screening ratio. Due to the non-linear nature of the relationship between ΔT_m and

K_d it is clear that that ranking compounds with small shifts in the melting temperature is open to considerable uncertainty even without considering the possibility that the ligand has some degree of affinity for partially folded forms of the protein or competition with solvents present in the buffer.

The inherent uncertainty associated with small ΔT_m could be addressed by increasing the ratio of ligand to protein. However, ligand solubility prevents this as a general solution. In this study the TDF assay was used as a first screen to group compounds into activity bands. It can be seen from Figure 6-9 that if $\Delta T_m > 4\ ^oC$ it is likely to equate to a ligand with $K_d < 50\ \mu M$ and for $\Delta T_m > 2\ ^oC$ it is likely to equate to a ligand with $K_d < 150\ \mu M$.

6.4.5 **Surface plasmon resonance**

SD1 was the lowest affinity ligand that was reliably detectable from the ligand test set.

Figure 6-10 SPR data for FKBP12 and SD1. Figure 6-10 (A) shows very fast association and dissociation rates (> 10 $\mu M^{-1}s^{-1}$) for SD1 at 25 ^{o}C at all ligand concentrations. The experiment was carried out from 0 to ~10x$K_{d(ITC)}$.

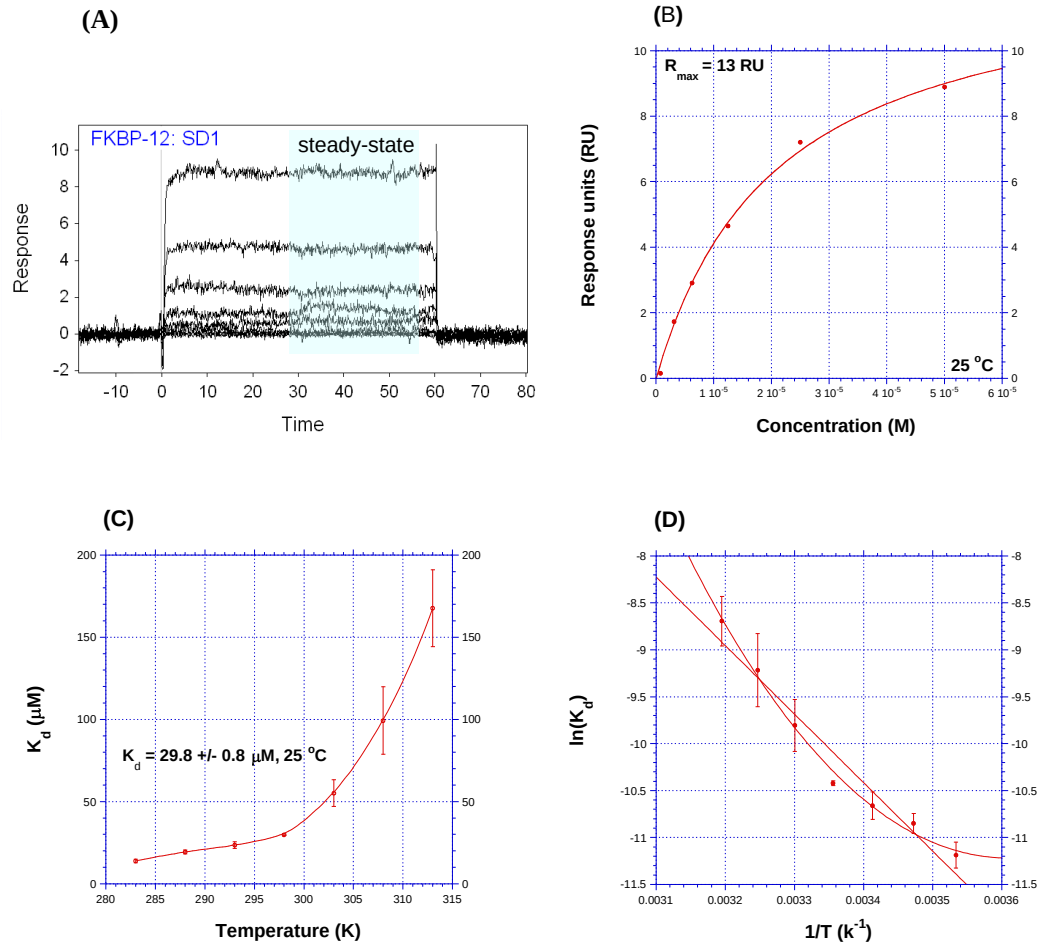


Figure 6-10 SPR FKBP12-SD1

(A) Representative SPR sensorgram showing response units for FKBP12 and SD1 from 0 to 50 µM at 25 ^{o}C. SD1 has a high association constant and a steady state was reached within the timescale of the experiment (shaded area). (B) Response units plotted against $[SD1]_M$ for representative data at 25 ^{o}C. Curve is the least-squared fit to the Langmuir model for 1:1, equation 26. (C) K_d determined from a 1:1 binding model, 10 to 40 ^{o}C (283 to 313 K , n = 3, standard errors, curve is added to aid visualisation. (D) Van't Hoff analysis of SPR data. Data was fit to equation 27 (straight line), the curved fit is the non-linear least squares fit to equation 28 that considers an enthalpic change with T (Kaleidagraph V4. 03, Synergy Software). This data was kindly with assistance Dr. Martin Wear (Edinburgh University).

It was not possible to derive kinetic data from the sensorgrams. The fast association rates at the high ligand concentrations meant that steady-state response levels were used for evaluating the affinity of the interaction, Figure 6-10 (B). R_{max} is not large for this system, ~ 13 RU. For comparison, 1.28 μM rapamycin (0.2 nM, 914 Da) gave R_{max} ~ 25 RU (saturating concentration). SD1 has affinity in the low micromolar range (6.8±1.2 μM, 25 ^{o}C, ITC) and a small mass, MW = 251.1 Da. Figure 6-10 (C) shows how K_d was varied with temperature; 10 to 40 ^{o}C in 5 ^{o}C intervals. There was a rapid dependence of K_d on T above 298 K (25 ^{o}C), there was also more variability in the data above this temperature.

SPR - FKBP1:SD1	***ΔH (kcal·mol^{-1})***	***TΔS (kcal·mol^{-1})***	***ΔC$_p$(kcal·K^{-1}·mol^{-1})***
Linear van't Hoff[a]	-14.5±1.5	-8.5±1.5	-
Non-linear van't Hoff[b]	-15.0±0.6	-8.9±0.63	-0.71±0.14

[a] equation 27, [b] equation 28

Table 6-5 van't Hoff analysis – SPR data FKBP12-SD1

SPR data from 10 to 40 ^{o}C was fit to a linear van't Hoff model ([a] equation 27) and a non-linear model ([b] equation 28) to examine the temperature dependence of the enthalpic change on ligand binding. Points plotted are mean values for n = 3, errors are from the fitting program.

Van't Hoff analysis was carried on the data first using a model where there was no dependence of the enthalpy of binding, ΔH, on temperature, equation 27. Non-linear van't Hoff analysis was also carried out considering temperature dependence for ΔH. Data was fit to equation 28 using least squares non-linear regression (Kaleidagraph V4. 03, Synergy Software). Due to the small temperature range and the considerable variability of the data above 300 K it is probably not sensible to put too much weight on the dissection of the free energy change into enthalpic and entropic components, Table 6-5. ITC predicted a small positive entropic contribution to binding at 25 ^{o}C (6.4.2). The modelling studies on SD1 in complex with FKBP12 (4.4.3) would suggest that the large negative ΔH on ligand binding as predicted by the van't Hoff analysis of the SPR data is unlikely, due to the small size of the ligand, limiting the possibility of van der Waals, contacts and the single putative hydrogen bond. The large negative ΔS is also unlikely as SD1 is a relatively hydrophobic molecule. It is thought that there is a favourable entropic change when two ordered waters are displaced from hydrogen bonding interaction with Tyr82 of FKBP12 (Connelly *et al.*, 1993; Burkhard *et al.*, 2000). The model described in 4.4.3 predicts a hydrogen bond between the carbonyl oxygen of SD1 and Tyr82 replacing the hydrogen bonds with water. A negative value for ΔC_p can be indicative of a hydrophobic interaction (Homans, 2007).This model gave a negative change in specific heat capacity, ΔC_p = -0.71 kcal·K^{-1}·mol^{-1}. A large negative value

for ΔC_p is more indicative of a hydrophobic interaction than the relative contributions of ΔH and ΔS to ΔG from van't Hoff analysis. When this value is used to extrapolate TDF data from T_m to 25 ^oC (6.4.4); calculated affinities are close to values measured by ITC for SD1.

6.4.6 **Screening the ligand test set- results summary**

ligand	***PPIase*** *% inhibition, 4 °C*		***ITC*** *K_d (μM), 25 °C*		***TDF*** *K_d (μM) 55 °C*		***ESI-MS*** *Micromass ZQ % in complex (8+) cV = 15 V*		*Literature K_d (μM)*	*Reference*
rapamycin	100±8.0[a]	1	0.007±0.005[z]	1	1.2±0.6	1	97.6±0.9	1	0.0002_{radio} to 0.008_{ITC}	(Bierer *et al.*, 1990; Connelly and Thomson, 1992; Wear and Walkinshaw, 2007)
FK506	97.8±8.5	2	-	2	1.3±0.6	2	83.1±0.5	2	0.0004_{radio}	(Bierer *et al.*, 1990; Connelly and Thomson, 1992)
KB2_61D	90.8 ±18.0	3	6.6±0.7[v]	3	18±3.5	3	74.5±9.5	3		section
SD1	69.7±9.3	4	6.8±1.2[w]	4	50±28	4	32.3±15.2	4	$\sim 9_{NMR}$	(Stebbins *et al.*, 2007)
SD12	27.2 ±9.7	6	-	-	133±68	6	protein unfolds	-	56_{NMR}	(Stebbins *et al.*, 2007)
SD2	40.8±9.9	5	-	-	151±93	7	18.7±4.0	5	95_{NMR}	(Stebbins *et al.*, 2007)
SD3	5.7±13.2	7	-	-	103±49	5	12.3±7.7	6	157_{NMR}	(Stebbins *et al.*, 2007)
SD4	-	-	-	-	-	-	No complex	-	221_{NMR}	(Stebbins *et al.*, 2007)
SD5	-	-	-	-	-	-	No complex	-	270_{NMR}	(Stebbins *et al.*, 2007)

Key: - experiment performed, no affinity data

[a] data kindly provided by Dr. Martin Wear (Edinburgh University).

Table 6-6 Biochemical and biophysical characterisation studies - screening the ligand test set

The ligand test set provided a useful series of compounds to compare screening assays. The results presented in the table are for screening at a single ligand concentration for the PPIase, TDF and ESI-MS assays. ITC data is included as a benchmark affinity measurement. Standard errors are shown for n=3 for all assays except the PPIase assay where n=9. The rank order is shown to the right of the affinity data. The solid red bar indicates the threshold affinity for the technique. Note assay protein/ligand screening ratios are similar but not identical in the different assays. The maximum ligand concentration for screening in PPIase, ITC and TDF is limited by solubility. High ligand concentrations are generally desirable for these assays to improve signal to noise ratio. For the ESI-MS assay ligand concentration is also limited by solubility but less so due to the methanol added to the analyte to improve desolvation. However, a high ligand concentration can lead to suppression of the protein signal and reduce signal to noise ratios for the species of interest or produce false positive from non-specific aggregation. **PPIase:** 100 nM FKBP12, 300 μM ligand, 0.6 $mg^{-1}ml^{-1}$ α-chymotrypsin, 120 μM Succ-Ala-Leu-Pro-Phe-*p*-nitroaniline, 14 mM LiCl, 1 % methanol (v/v), 3 % 2,2,2-trifluoroethanol (v/v), PBS, pH 8, 4 °C. [z] **ITC:** (FKBP12-rapamycin): 34.9 μM FKBP12 (syringe), 0.984 μM rapamycin (cell), 50 mM sodium phosphate, 50 mM sodium chloride, pH 7, 25 °C. [v] **ITC**: (FKBP12-KB2_61D): 15 μM FKBP12 (cell), 300 μM KB2_61D (syringe), 50 mM sodium phosphate, 50 mM sodium chloride, pH 7, 25 °C. [w] **ITC:** (FKBP12-SD1): 13.74 μM FKBP12 (cell), 300 μM SD1 (syringe), 50 mM sodium phosphate, 50 mM sodium chloride, pH 7, 25 °C. **TDF:** 10 μM FKBP12, 300 μM ligand, 10x SYPRO, 50 mM ammonium acetate, 0.2 % $DMSO^{final}$, 1 % $methanol^{final}$. **ESI-MS:** 10 μM FKBP12, 100 μM ligand, 19 mM ammonium acetate, pH 6.8.

Screening results need to be considered in the context of the time for the assay, reagent consumption and cost. This information is presented in Table 6-7. For the purposes of comparison a screen of 96 compounds is considered and a ligand of 350 Da.

Technique (instrument)	***N***	***General screening concentrations***	***Protein requirement /96 compounds***	***Screening compound (MW 350 Da)***	***Time 96 compounds***	***Minimum affinity detected (μM)***
PPIase assay (Perkin-Elmer Lambda 20)	9	Coupled assay (100 nM FKBP12, 120 μM substrate) 300 μM ligand	1.5 mg	1.3 mg	6 weeks[w]	157
TDF (Bio-Rad iCycler IQ™, iQ5 real time detection)	3	10 μM FKBP12 300 μM ligand	1.1 mg	21 μg	2 days[w]	157
ESI-MS (Micromass ZQ)	3	10 μM FKBP12 100 μM ligand	3.4 mg	8.8 μg	5 days[w]	157
SPR (Biacore T100)	3	Concentration of protein NA 50 μM ligand	~4 μg	2 μg	2 days	6.8
ITC (MicroCal auto-VPITC)	3	15 μM FKBP12 300 μM ligand	125 mg	0.5 mg	12 weeks	6.8

Table 6-7 Comparison of Biochemical and Biophysical screening techniques – reagent consumption and timing

The table summarises the general screening conditions and the protein consumption and time to assay 96 compounds. The ligand consumption is calculated per ligand, a typical MW of 350 Da was chosen for the calculation. The use of standard controls has been included and repeats as detailed for each assay. [w] indicates that the assay is not automated for a 96 plate and can only be carried out over the working day without robotic assistance.

6.5 Discussion and conclusions

Choice of screening approach required balancing throughput, sensitivity and the richness of thermodynamic data. A general comparison of advantages and disadvantages of each method are listed in Table 6-8. It is useful to consider these in tandem with Table 6-7 which details assay times, reagent consumption and results for the ligand test set. It is clear that no one method can provide all the information required within in a reasonable time scale or cost. Each method delivers an apparent K_d that must be considered with the inherent assumptions made in the model and the possibility of false positives. Compounds from the ligand test set were used to compare different methods and lead to screening strategy that could be used to test the hits from virtual screening programs. A proposed screening strategy is shown in Figure 6-11.

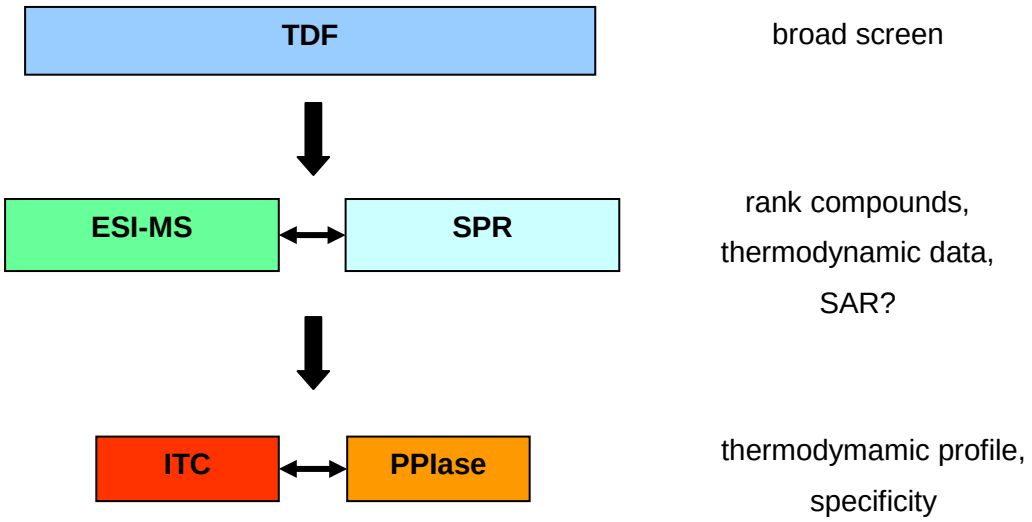


Figure 6-11 Screening strategy

The strategy divides the assays into 3 tiers based on throughput and information. TDF and SPR provide the highest throughput assays (~2 days for 96 compounds). TDF was chosen as the first step due to the considerably lower cost of consumables and the increased sensitivity for low affinity ligands in the protocols tested. ESI-MS and SPR form the second tier in the screen, as they provide a fast route to thermodynamic information.

Technique (instrument)	*T (°C)*	*Advantages*	*Disadvantages*
PPIase assay (Perkin-Elmer Lambda 20)	4	• Rank compounds by affinity • Determines specificity for the active site	• Low throughput: ~3 compounds/day, many repeats required • Coloured and fluorescent compounds a problem • Coupled assay • Manual dexterity needed • High substrate/ screening compound consumption
TDF (Bio-Rad iCycler IQ™, iQ5 real time detection)	~55	• Rank compounds by affinity • High throughput (parellised) • Low DMSO concentrations can be accounted for • Automated data capture	• Ligand solubility a problem • Coloured and fluorescent compounds a problem • Extrapolating K_d to 25 °C requires knowledge of ΔH and ΔC_p or bracketing
ESI-MS (Micromass ZQ)	30?	• Rank compounds by affinity • Monitors protein folding/contamination • Confirms mass of protein/ligand • Fluorescent and coloured compounds ok • 10 % methanol in analyte aids compound solubility • Possibility of compound mixtures • Does ESI-MS estimate ΔH not ΔG?	• No DMSO • Are hydrophobic ligands poorly represented in ranked lists? • Real temperature hard to measure
SPR (Biacore T100)	4-45	• Rank compounds by affinity • High throughput • Kinetic data • Automated data capture • Low reagent consumption	• Expensive chips • Solvents can cause noise in data • No real parallisation on T100 • High capital outlay for instrument
ITC (MicroCal auto-VPITC)	2-80	• Highest quality thermodynamic data, direct measurement of ΔH • Dissection of ΔH and ΔS from ΔG in a single experiment • Automated data capture	• Low throughput: ~6 compounds/day • Fitting v. sensitive to reagent concentration • Correction for heats of dilution • Reagent consumption • High capital outlay for instrument

Table 6-8 Comparison of Biochemical and Biophysical screening techniques – general considerations.

The table summarises the advantages and disadvantages for each assay for screening a single ligand concentration against FKBP12. The temperature range of each instrument is also included. In the case of ESI-MS, temperature can be increased and decreased but values are hard to quantify.

Estimating K_d from TDF is limited by the need to extrapolate data to 25 °C using ΔH_L and ΔC_p. Information that is not available before a detailed ITC study. The enthalpic contribution to binding can be quite variable for structurally diverse ligands of similar size. This means

that the ranking position of a ligand at the protein melting temperature can be quite different from that at 25 ^{o}C. This study suggests that the best approach is to use TDF to select a subset of the virtual screening hits to be tested by the other methods. A cut-off ΔT_m of ~2 ^{o}C should pick up ligands of K_d < 150 μM. Chapter 4 describes how ESI-MS can be used to dissect the enthalpic contribution to binding. SPR has the potential to be carried out at different temperatures and also to provide kinetic information. ESI-MS and SPR can provide richer information for a SAR study than TDF. In addition, these methods provide information on the stoichiometry of the interaction. Useful information, particularly in fragment screens where ligands might be non-specifically "sticky".

The throughput of the PPIase assay was not high enough for the assay to be selected as the first stage in a medium throughput screening strategy. This was in part due to the need for a high number of repeats to gain a meaningful rank of ligand by affinity. The principle benefit of the technique is that it confirms specificity for the active site of FKBP12 by a biochemical method. We would propose that it is used as an element of the final tier of screening, alongside ITC.

ITC measures the enthalpy of complex formation. Most other techniques rely upon calculation of K_d from the measurement of a correlated physical parameter by use of a model with varying degrees of complexity. The rich thermodynamic data needs to be taken in the context of the greater demand for protein and ligand and a longer experimental time. The experiments designed to increase the throughput of ITC by reducing the c-value of the experiment were disappointing. The method performed well for ligands with a K_d < 10 μM but failed to pick up the low affinity ligands. In the screening strategy proposed, ITC is placed in the final tier providing "gold standard" thermodynamic information.

7 Chapter 7 Screening for Inhibitors of FKBP12

7.1 Introduction

Compounds from four libraries were screened against FKBP12; EDULISS 2 (3.8 M unique compounds from commercial and academic sources (Hsin, 2009) ; a subset of the Specs catalogue (15000 compounds. www.specs.net; Jan 2009); a fragment library (96 compounds selected from The Maybridge Ro3 collection, UK) and the KB series (a library designed to target immunophilins).

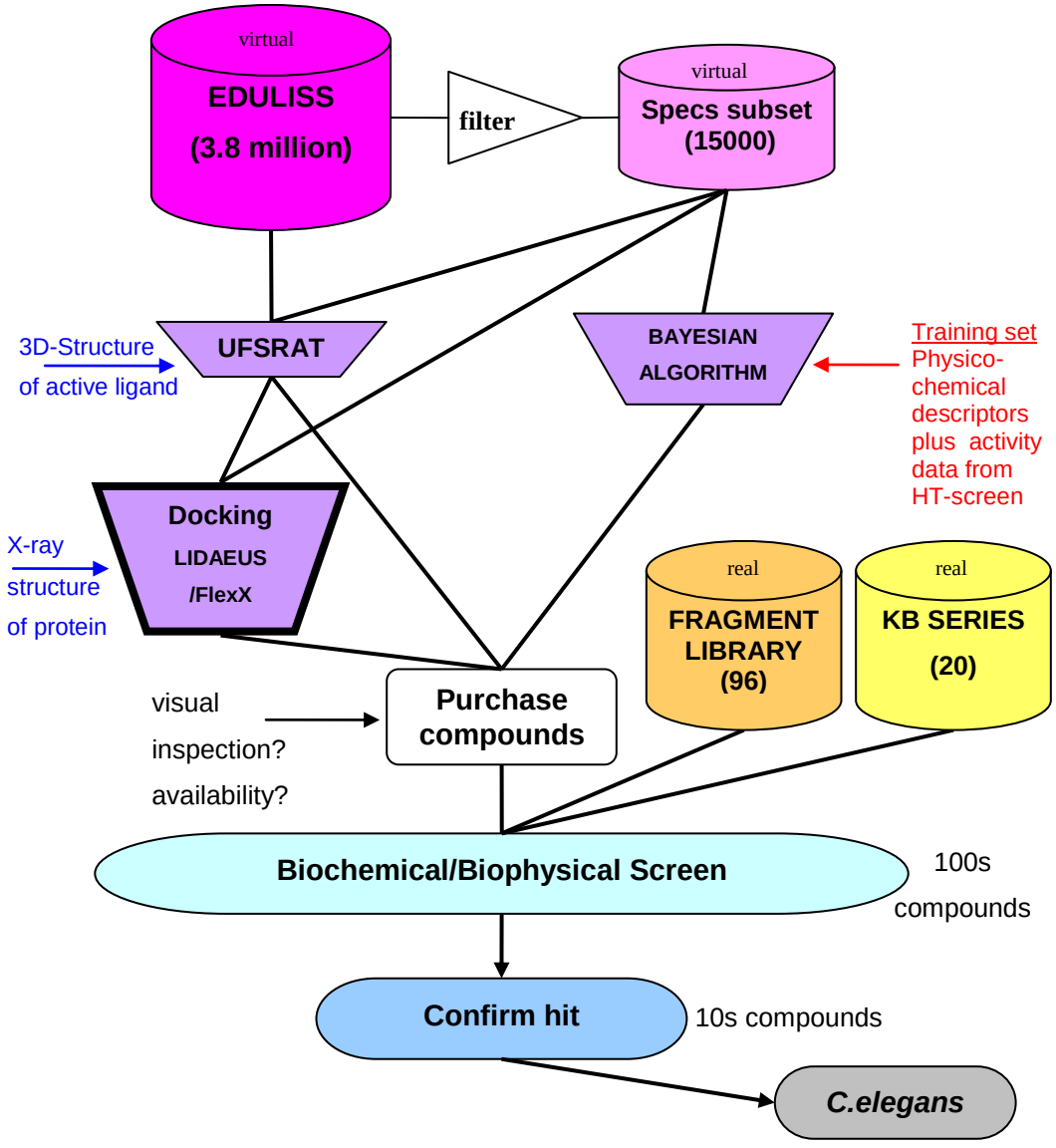


Figure 7-1 Screening space

Compounds followed two paths to "wet" laboratory testing: from virtual collections via screening programs and from two in-house "real" compound collections. The "drum" library icons are labelled with the number of compounds in the library. The horizontal arrows link to the structural data input.

Figure 7-1 shows an overview of the path of compounds from a library through screening techniques from virtual to experimental. It is useful to make a distinction between the virtual libraries and the "real" in-house compound collections. Screening assays are discussed in detail in Chapter 6.

The large libraries of virtual compounds were initially screened with *in silico* or virtual screening approaches. The aim was to score the compounds in such a way that the highest ranking compounds in a list had a higher percentage of active compounds than a random selection. The final number selected was dependent on both the cost of compounds and the throughput/cost of the assay. Compounds from the virtual libraries were all passed through filters to select for drug-likeness and one or more virtual screening programs. The virtual screening strategies all had an input of structural data specific to the protein target. Each method took the structural information in a different form and then applied it in a unique way. The virtual screening methods are described in more detail in the following sections. UFSRAT (Steven Shave, Edinburgh University) and the Bayesian clustering algorithm (Angelopoulos *et al.*, 2009) are ligand based methods. LIDAEUS (Wu *et al.*, 2003) and FlexX (Rarey *et al.*, 1996) are well established macromolecular docking programs that use protein and ligand structure to make binding energy predictions.

UFSRAT compared the 3D-structure and electrostatic properties of a library compound to a ligand of known activity and gave it a score based on similarity. The Bayesian clustering algorithm used physicochemical/structural descriptors and experimental activity data from a training set (known actives and inactives of FKBP12) to assign an activity probability for a specific library compound. The docking programs LIDAEUS and FlexX used an x-ray structure of FKBP12 to predict the enthalpy of binding of a putative ligand.

The selection of the in-house compounds collections were also guided by structural considerations. However, the principles of compound selection were very different. The fragment library was not specifically designed to target FKBP12. The aim was find novel scaffolds (chemotypes) that bound to FKBP12 (Schuffenhauer *et al.*, 2006; Ertl *et al.*, 2006). The fragment based approach to drug design has two central tenets. Firstly, that is more efficient to probe chemical space by using fragment-like compounds. Secondly, that the free-energy contribution to binding per atom, the ligand efficiency, is greater for a weakly binding fragment than for the larger molecule typically seen in a HT-screen (Kuntz *et al.*, 1999; Hesterkamp and Whittaker, 2008). It is implicit that low affinity fragments are more

tractable to lead optimisation than low affinity larger molecules (Ciulli and Abell, 2007; Reynolds *et al.*, 2008). In contrast, the KB series of compounds was designed and synthesised by Dr. Kevin Bailey as a library to specifically target two structurally different immunophilins; FKBP12 and human cyclophilin A (section 1.11.1 describes the immunophilin proteins). Structure activity relationships were derived from a literature review of known FKBP12 and cyclophilin inhibitors, section 1.11.6. This method was not designed to select novel scaffolds but to further the understanding of these enzymes.

7.2 LIDAEUS

7.2.1 Ligands discovered using the program LIDAEUS

The bulk of virtual screening employed Edinburgh University's in-house docking program LIDAEUS (**LI**gand **D**esign **A**t **E**dinburgh **U**niversity). LIDAEUS has found novel inhibitors for 3 protein targets: cyclin dependent kinase 2 (CDK2), polo-like kinase (PLK1) and cyclophilin A (hcypA). The best in series compound inhibiting CDK2 was CYC3, with an IC_{50} of 2.2 μM (Wu *et al.*, 2003). Cyclapolin 9 binds in the ATP-binding site of Polo-like kinase (PLK1) with an IC_{50} of 20 nM (McInnes *et al.*, 2006). The inhibitor series for hcypA are interesting because they are built upon the small fragment-like molecule dimedone (Yang *et al.*, 2007). Dimedone was picked up as hit in a virtual screen and subsequently crystallised in complex with hcypA. SANDOCK, a precursor to the program LIDAEUS written by the same principle authors (Taylor and Walkinshaw), discovered a diverse set of ligands for the protein FKBP12 (Burkhard *et al.*, 1999). The structure of the steroid found to bind to FKBP12 is shown in Table 1-4.

7.2.2 Docking and scoring using the LIDAEUS pipeline

Docking using the program LIDAEUS is divided into a series of discrete steps or modules. The modules comprising the LIDAEUS pipeline are illustrated in Figure 7-2.

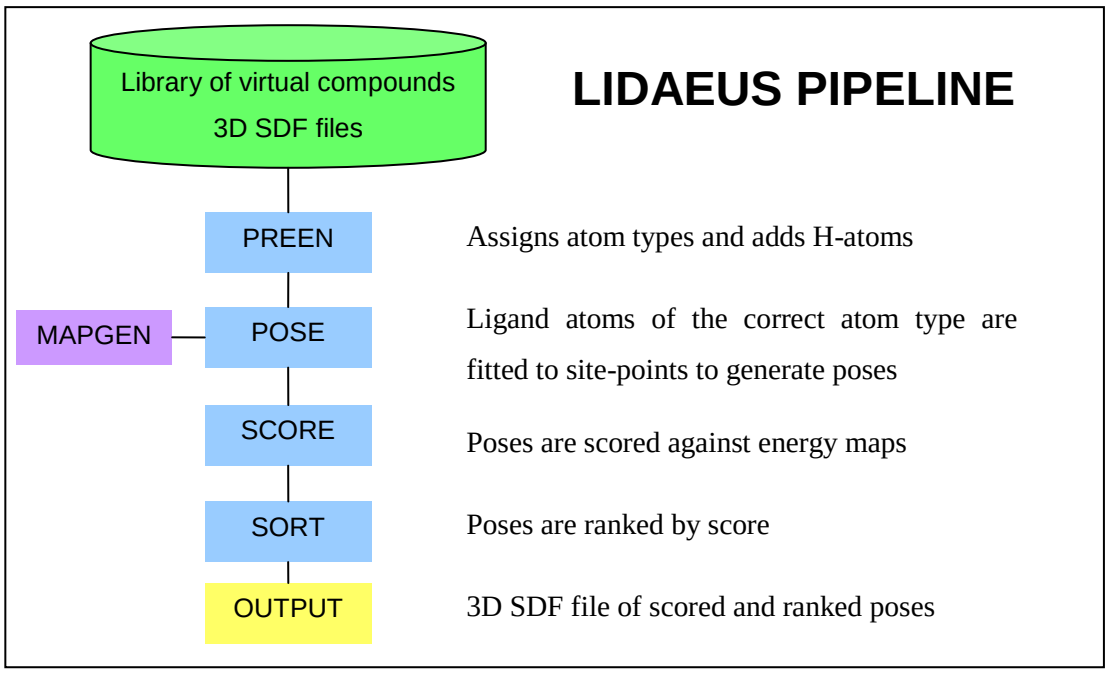


Figure 7-2 The LIDAEUS docking pipeline

7.2.3 Site-point generation with MAPGEN

LIDAEUS uses high quality crystallolographic data, of a protein in complex with a ligand, to generate a 3D cubic grid of site-points in the binding pocket of the target protein. The binding pocket is defined by inputting the coordinates of a molecule in the active site of the protein in the form of a SDF file. This molecule is used as series of reference points to define the dimensions and placement of the grid; it need have no physiological relevance and it is possible to define the volume around the ligand to explore (for example, 4 Å beyond ligand coordinates). The program MAPGEN was used to generate a set of energy maps to guide the program in the assignment of the classification (type) of site-points to the each position in the grid and to score ligand poses. Site-points are classified hydrogen-bond donor (blue in visualisations), hydrogen-bond acceptor (red) and hydrophobic (white). Energy maps are calculated that describe hydrophobic, extremely hydrophobic, hydrogen bond donor and hydrogen bond acceptor potentials for each grid point. Each grid point may have only one type of site-point assigned. There is a user definable energy cut-off value for the minimum energy required to assign a site-point to a particular type (colour). If a more than one site-point criterion is met the potential energy for each map is compared and a series of rules used to assign the colour of the site-point.

The site points generated for FKBP12 were clustered in regions of the binding pocket where non-covalent interactions were most likely. For example in the FKBP12 binding pocket there were hydrophobic site-points assigned near the Trp59 residue that forms the base of the pocket (Figure 1-5 and Figure 7-3). Similarly, hydrogen bond acceptor site-points were assigned near the main chain N of Ile56 that donates an H atom in an H-bond interaction with rapamycin (ligand atoms that act as H-bond acceptors are docked onto acceptor site-points).

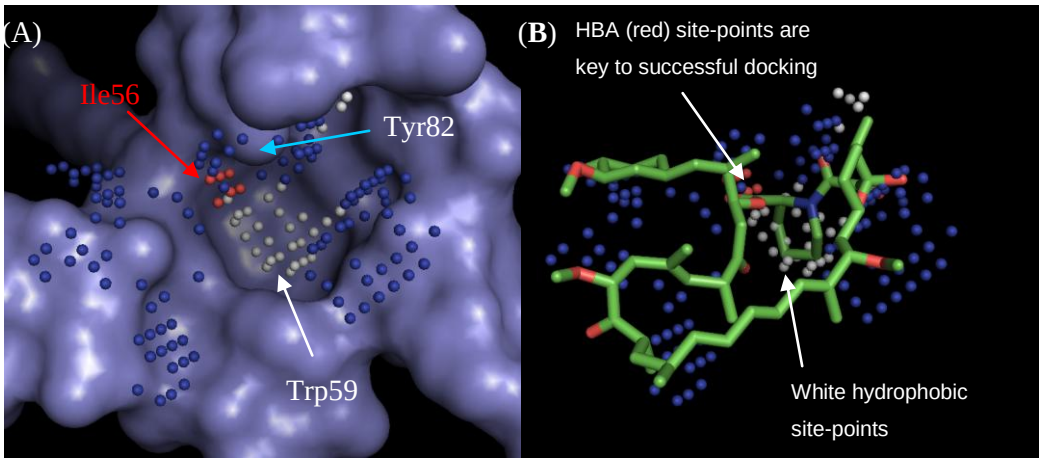


Figure 7-3 Site-points for FKBP12

(A) Shows site points in the active site of FKBP12. Blue: hydrogen bond donor (HBD); Red: hydrogen bond acceptor (HBA); white: hydrophobic. A Connelly surface has been added to FKBP12 to aid visualisation of the active site using the program PyMOL (V1.0r2, DeLano Scientific). (B) The structure of rapamycin in complex with FKBP12 in the same frame of reference as the site-points.

7.2.4 PREEN

The PREEN module assigns atom types and charges to each atom of the molecules (ligands) to be docked, using the coordinates and connectivity information in the input SDF file. For example a sp3 hybridised carbon atom is distinguished from a sp2 and an atom that can act as a hydrogen bond donor from one that can act as a hydrogen bond acceptor. This is necessary for fitting atoms of the correct type onto site-points during posing of the ligands in the active site. Hydrogen atoms are added to the SDF file.

7.2.5 **POSE**

POSE geometrically matches pairs of ligand atoms to pairs of site points to generate poses. The user may define what types of atoms are fitted to a particular type of site-point. For example, under standard operating parameters ligand atoms that act as H-bond acceptors are docked onto acceptor site-points and atoms that act as donors onto donor site-points. The user may also define the number of matches between ligand atoms and site points. The greater the number of matches the longer the run time and the fewer poses generated. The resolution setting in LIDAEUS determines the allowable difference in distances between ligand atoms and site points. For each subsequent match this requirement is relaxed. The higher the resolution set the greater the number of poses generated and the longer the run time. A resolution of 0.04 Å was the default setting in this study. Default settings were used for atom typing, matching and energy cut-offs. Alternative values for energy cut-offs were tested using crystallographic data as a reference structure to compare with docking results. These alternative cut-offs did not improve RMSD between docking results and crystallographic poses and were rejected.

7.2.6 **SCORE and SORT**

The SCORE function in LIDAEUS calculates an approximation to the free energy of binding for each pose. Ligands then undergo a rigid body force field energy minimisation to optimise the pose and predicted binding affinity, ligands are then are sorted by SCORE. The more negative the score the more favourable the interaction. LIDAEUS uses an empirical scoring function that combines the energy scores for each atom in a ligand pose from the energy maps generated in MAPGEN. A separate score is outputted for each map and also a combined score which may be a simple summation or weighted (user defined) in favour of a particular map. It is possible to define the number of poses outputted and/or use a score cut-off to reject poorly scoring poses.

The program has been adapted to run on the Blue Gene/l (IBM, 2005) supercomputer (Shave *et al.*, 2008). This allows the program to run on up to 2048 parallel processors and vastly reduces the run time for docking. This has increased the potential for docking large, diverse sets of compounds and also analysing hits after changes to docking template or program parameters. Using the LIDAEUS suite of programs the experimenter can influence site-point generation, and therefore docking near to key residues (“hot spots” identified from x-ray structures of known tight binders).

7.2.7 Choice of docking template influences LIDAEUS docking

When structures of FKBP12 in complex with inhibitors were compared with the apo structure it is clear that there are small changes on ligand binding, principally in the regions of the 80s and 40s loops (Figure 1-8 shows an alignment of 12 x-ray structures). For this reason a structure of FKBP12 in complex with a ligand was considered the most suitable conformation (template) for docking experiments. Waters were not included in the docking template as all the x-ray structures deposited in the PDB show the 3 waters, seen in the active site of the apo structure, to be displaced on ligand binding (Figure 1-9). A test set of ligands identified in crystallographic studies were docked to compare performance between templates. The RMS deviation of the docked pose from the crystallographic pose was used as the metric for docking success. Enrichment rates were also compared by plotting ROC (receiver operating characteristic) curves (Verdonk *et al.*, 2004; Konstantinou-Kirtay *et al.*, 2007). The most successful template proved to be 2DG3.pdb (FKBP12 in complex with rapamycin, (Fulton *et al.*, 2003)). LIDAEUS docks rapamycin (x-ray conformation minimised in TRIPOS) with an RMSD of 1.02Å using the 2DG3.pdb template. Examining the docking results and the site-points generated for the different templates showed key differences that were important for good enrichment and low RMSD. Hydrogen-bond acceptor site-points associated with the main chain N of Ile56 are the key to successful docking. H-bonding interactions between Ile56 and the ligand are seen in all known x-ray structures of FKBP12 small molecule complexes (Table 1-3). Some templates have subtle differences in structure that mean that Tyr82 (a possible H-bond donor) dominated calculated potential energy contributions at grid points between Ile56 and Tyr82 and MAPGEN assigned site-points as hydrogen bond donor in these positions. 2DG3.pdb was used as the docking template in subsequent docking experiments. 2DG3.pdb was also used as the template for docking with FlexX.

7.3 FlexX

FlexX uses a fragment-based approach to docking (Rarey *et al.*, 1996). The initial step in docking is to break the ligand up into fragments; the principle fragment is known as the base fragment. A relatively rigid base fragment is chosen with fewer low energy conformations than alternative deconstructions of the molecule. Weight is also given to selecting base fragments that make more non-covalent interactions with the protein surface. In the same way that relaxing the resolution setting in LIDAEUS increases the possible number of poses and therefore docking time, small base fragments in FlexX extend computational time. Once chosen, the base fragment is considered a rigid body and triangles of atoms in the molecule

are geometrically fitted to a description of the protein surface. Potential ligand poses are clustered by considering the root mean squared deviation between poses. Members of a close cluster are mathematically combined into a single solution. Solutions are then compared and ranked in terms of the possible interaction energy of the fragment with the protein. The ligand is then incrementally grown from the best solution at each stage of incremental growth. The program makes use of tables of low energy torsion angles derived from the Cambridge structural database.

Docking in FlexX was performed using the crystal structure of rapamycin in complex with FKBP12 (2DG3.pdb) as the template. 6.5 Å was explored around the template ligand and default charges used for ligand and protein. 3D SDF files of ligands were converted to mol2 files prior to docking using the facility in the FlexX suite of programs. The top scoring pose was saved for each ligand using the standard FlexX scoring function. This is an empirical scoring function based on the work of Boehm and a test set of experimental data of binding constants for 82 protein ligand complexes (Klebe and Boehm, 1997; Boehm, 1998).

7.4 Validating docking results

It is widely acknowledged that fast empirical scoring functions in docking programs such as FlexX and LIDAEUS do not predict the free energy of binding of a ligand with a high degree of accuracy (Perola *et al.*, 2004; Feher, 2006; Konstantinou-Kirtay *et al.*, 2007; Waszkowycz, 2008). Docking results are generally validated by two metrics. Firstly, the ability of the algorithm to pick (score most highly) the correct pose of the ligand in the active site. Secondly, a metric that measure enrichment as a function of ranked score. Root mean square deviation (RMSD) is routinely used to compare the highest scoring docking pose of a ligand with the crystallographic pose. A cut-off of 2 Å is often used to define a "good" docking pose (Perola *et al.*, 2004). Receiver operating characteristic curves (ROC curves) can be used to graphically illustrate the proportion of hits as a function of score (Konstantinou-Kirtay *et al.*, 2007).

7.5 Ligand-based virtual screening

7.5.1 Chemical similarity

The concept of chemical similarity is useful but also one which is hard to define. The central assumption of the similarity principle is that molecules with a similar structure have similar physicochemical properties and similar biological activity (Bender and Glen, 2004). There is no such thing as absolute similarity; rather molecule A can be considered similar to molecule

B relative to some measurable or definable element. This element could be a specific functional group, an atom path/graph within a molecule or a descriptor defining some global property of the molecule. It is also hard to define similar biological activity. For example, when conducting a similarity search in EDULISS 2 compound SD1 is retrieved as a close neighbour of compound SD2 by the program UFSRAT (top 0.1 % in 5.3 million compounds). SD1 has a Tanimoto similarity coefficient > 0. 9 when compared to SD2 using a fingerprint based method. The difference in structure between SD1 and SD2 is a substitution of –S– for $–SO_2–$; this resulted in a 10 fold loss of activity. Replacing –S– with –O– reduced activity 50 fold (Stebbins *et al.*, 2007). This result could rightly be considered a triumph of the program UFSRAT and the similarity principle in purely structural terms. However, in most experimental assays it is extremely difficult to pick up an inhibitor in the 100s of μM range at standard screening concentrations. Where does similarity in biological activity end? Is an 8 μM inhibitor similar to a 100 μM inhibitor or a 400 μM one?

Whatever the complexities in comparing the structural features of one molecule with another, the concept of chemical similarity is frequently utilised in drug discovery. The ability to assess the diversity of a large chemical library (real or virtual) is very useful, particularly when deciding on the compounds to include in an HTS assay. In the context of virtual screening it is very useful to assess chemical similarity in the "post processing" stage of docking experiments. "Post processing" is the slightly ambiguous term that is used to describe the analysis of docking results. At an early stage of this project it became apparent that when the top 1000 hits of a LIDAEUS run were visually inspected there were chemical series in the hits. Tools were used to cluster the hits from a docking run by chemical similarity. This method allowed the experimenter to sample the chemical space of the hits more efficiently when only a small number of compounds could be purchased and tested. Clustering compounds by chemical similarity places less emphasis on the detailed ranking provided by a scoring function. This tool was also very useful when looking at SAR in experimental data.

Chemoinformatic techniques that are used in similarity studies to define a molecule in terms of a 2D fingerprint can also be used to relate the complexity of molecules to bioactivity (Schuffenhauer *et al.*, 2006). Similar techniques have been used to compare the variety of ring scaffolds in screening collections and known drugs (Ertl *et al.*, 2006).

7.5.2 Clustering compounds by chemical similarity

7.5.3 Molecular fingerprints and similarity metrics

Molecular fingerprints (2D string representations of a molecule) were used to make pair wise comparisons of molecules in a compound collection. A binary fingerprint is a string of 1s and 0s describing the features of a molecule. If a feature is present the bit is set to 1 and if it is absent to 0. The feature may be a chemical functional group; an atom count; a substructure within the molecule; or a path length between a specific pair of atoms/functional groups/ring systems/pharmacophores Commonly used fingerprints are those from MDL Information Systems Inc. (Durant *et al.*, 2002) and Barnard Chemical Information Ltd. (BCI fingerprints, (Barnard and Downs, 1997)). 2D hashed fingerprints are also commonly used; they differ from binary fingerprints in they do not use a predefined dictionary of fragments. Common commercially available hashed fingerprint algorithms are Daylight fingerprints (www.daylight.com; Jan 2009) and the Unity fingerprint. When comparing two fingerprints it has been found that some methods are more sensitive to changes in chemical linkers than other methods and this must be borne in mind when viewing results (Stahl *et al.*, 2005). 2D fingerprints can be calculated from the coordinates and connectivity table in a 2D structure file or from SMILES strings (Weininger, 1988). In this study fingerprints were encoded using a proprietary method in the program Accord for Excel (Accelrys).

Two molecules can be assessed for similarity by comparing their fingerprints using a distance or similarity coefficient. In this study the Tanimoto coefficient similarity coefficient was used. The Tanimoto coefficient has been shown to perform well with chemical fingerprints (Willett, 2006). Two identical molecules have a Tanimoto similarity coefficient of 1 and two completely different molecules a coefficient of 0. The Tanimoto coefficient for two molecules A and B is defined by equation 29.

$$T_{A,B} = \frac{N_{AB}}{(N_A + N_B) - N_{AB}} \qquad \textbf{Equation 29}$$

Where: N_A is the number of bits set (1) in string A, N_B is the number of bits set to 1 in string B and N_{AB} the number of common bits set to 1.

Defining rules for cut-offs of Tanimoto coefficients depends on the fingerprint employed and is rather open to the interpretation of an individual chemist. As a rule of thumb, the Accord

fingerprint gives a Tanimoto coefficient > 0.7 for two very similar molecules and > 0.5 for two broadly similar molecules. For a given molecule of interest, ordering the remaining molecules in a set by Tanimoto coefficient can be a useful tool to pick close analogues in a large library. Fingerprints may be pre-generated and stored in a relational database. It is worth noting that the Accord fingerprints are generated from 2D SDF and two molecules deemed very similar by this method could have quite different 3D shape due to the replacement of a linker.

If a pair wise comparison is made of all the molecules in a set, it is possible to generate a symmetrical similarity matrix of dimensions NxN. (where N is the number of molecules in the set). The diagonal of this matrix is made up of 1s for matrix using the Tanimoto similarity coefficient (the diagonal represents the molecule being compared with itself). The similarity matrix was used as the input to a clustering algorithm to group the molecules into clusters and produce a dendrogram displaying chemical similarity.

7.5.4 **Clustering algorithms**

In this study the simple hierarchical agglomerative clustering algorithm UPGMA (unweighted pair-group method with arithmetic mean) was used to cluster the molecules in a set (MVSP V3.13, Kovach Computing Services). Data was input in the form of a similarity matrix. Using this methodology the output in the form of a dendrogram was useful for a visual overview of the diversity of a set of 100 compounds or fewer. Above this number the diagram was difficult to interpret visually. Figure 7-18 shows the UPGMA clustering of the KB2 series of compounds.

7.5.5 **UFSRAT**

UFSRAT (**U**ltra **F**ast **S**hape **R**ecognition and **A**tom **T**yping) is a virtual screening program developed by Steven Shave (Edinburgh University) that utilises molecular similarity. It is a database mining program that selects compounds based on a comparison of shape and electrostatic properties. UFSRAT was used to select compounds from the EDULISS 2 database with similarity to known inhibitors of FKBP12. Data is input in the form of 3D SDF files and output in the form of a list of compounds ranked by similarity score. The similarity scores from UFSRAT are normalised so that two identical molecules received a similarity score of 1 and two completely unlike molecules a score of 0.

UFSRAT encodes the shape of a molecule using metrics that describe the statistical distribution of Euclidean distance of individual atoms from a reference point. There are 16 points of reference in total: 4 for each atom type (hydrogen bond donor, hydrogen bond acceptor and hydrophobic) and 4 for all the atoms in the molecule. This method of describing molecular shape is an extension of the work of Ballester and Richards (Ballester and Richards, 2007). UFSRAT can be distinguished from Ballester and Richard's program as it considers atom types and therefore encodes a description of both overall shape and electrostatic properties. Atoms are classified as hydrogen bond donor, hydrogen bond acceptor or hydrophobic using similar criteria to those applied in the PREEN module of LIDAEUS. Each of the set of four points of reference comprises: the geometric molecular center (centroid), the closest atom and farthest atom to the centroid and the farthest atom from the farthest atom. Each distribution of distances is described using the standard statistical moments of mean, variance and skew. Each point of the 16 reference points has 3 descriptors giving 48 in total.

The reductionist approach of UFSRAT, in which 3D information is encoded in a 1D vector, allows for fast computational comparison of compounds. The program can compare 1million compounds per hour with a query molecule using pre-generated 3D SDF files. In this mode descriptors are generated on the fly. Using precalculated descriptors 1 million compounds can be compared to the query in 3 s.

UFSRAT has several features that distinguish it from the many similarity searches based on 2D molecular fingerprints. Fingerprints such as MDL and DAYLIGHT encode specific features within a molecule. A Tanimoto coefficient compares binary fingerprints by comparison of specific bits/features within a molecule. To state the obvious, a similarity search using this method retrieves molecules with similar features. As UFSRAT is not prescriptive of bond order or functionality the program has the potential to "scaffold hop". "Scaffold hopping" is the retrieval of molecules with similar shape and properties but with a different carbon skeleton (Boehm *et al.*, 2004). The emphasis of the algorithm on shape is of particular note. An inhibitor must have a complimentary shape to the binding pocket of the protein. A change to a linker between, for example, two ring systems can profoundly alter the overall shape of a molecule but might not be given weight in a 2D fingerprint. Such a change might make the molecule unable to exploit a binding pocket (the van der Waals potential describes the strong repulsive force when atoms come too close to one another). The emphasis of shape is perhaps both the strength and the weakness of UFSRAT.

Comparison by UFSRAT is sensitive to the conformation of the molecules in question. This is not immediately apparent when a known active is retrieved from a database, as the biological conformation of the molecule need not be known. In this case it is only important that the two conformations are consistent. In this study a minimum energy conformation in a vacuum generated by the program CONCORD (TRIPOS) was used for each molecule in the database. One could argue that set of active compounds retrieved using this algorithm might be richer when searching for more rigid molecules (fewer rotatable bonds). This would mean that the search compound had a higher probability of being the correct conformer.

UFSRAT was used to select two series of compounds in this study.

7.5.6 Using Bayesian machine learning to predict affinity

The virtual screening technique described (very briefly) in this section was developed by Dr. Nicos Angelopoulos (Angelopoulos *et al.*, 2009). Dr. Angelopoulos applied a Bayesian averaging technique to learn classification trees from high-dimensionality descriptors and affinity data and used these to predict relative ligand binding affinities. Compounds predicted to have a high probability of activity against FKBP12 were selected for testing.

Experimental affinity data was taken from the San Diego NMR high throughput screen of small molecules against FKBP12 (http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=608, PubChem BioAssay database: SDCCG-A014-FKBP12-NMR). This data is referred to as "the training set". For the purpose of the model, training set molecules were classified as active or inactive. Molecules were considered active if they had an estimated $K_d \leq 500$ μM. Long vectors of descriptor values were built for each molecule in "the test set" and the screening collection of molecules from the descriptors deposited in the EDULISS 2 database. A Markov chain Monte Carlo (MCMC) algorithm was used to construct a long chain of possible classification trees for the screening collection. In a classification tree, a specific magnitude of a descriptor determines the partitioning of the molecules at each internal node (forking of a branch). Molecules with a descriptor value smaller than a certain magnitude follow one edge (branch) and those larger the other edge (branch), Figure 7-4 Example of an internal node in a classification tree.

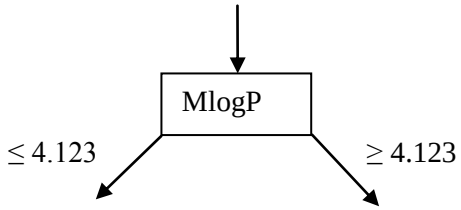


Figure 7-4 Example of an internal node in a classification tree

The partitioning of molecules at this node is determined by a value of MlogP smaller or larger than 4.123.

The tree ends in terminal nodes (leaves), each one representing a specific molecule in the library. Each leaf in the model gives a probability value of activity (0-1) to the molecule. This probability is derived from the proportion of actives to inactives in "the training set". The whole chain of trees is used to give an average probability of activity for each molecule in the screening collection. The average probabilities of activity for each molecule are used to create a ranked list.

The Specs green collection was screened using the Bayesian machine learning method (7.7.4).

7.6 Choosing drug-like compounds

A question that arose in this study was whether it was better to pass compounds through a drug-likeness filter prior to virtual screening and immediately exclude compounds that failed or, to filter compounds after virtual screening. The approach taken depended on the time taken in the individual screening steps. The majority of the compounds in EDULISS 2 pass Lipinski's criteria. LIDAEUS is an extremely fast docking program; the parellel version of LIDAEUS running on BlueGene (IBM) can dock 5 million compounds in a matter of hours (Shave *et al.*, 2008). In contrast docking 15,000 compounds with FlexX took over a week on a multiprocessor server. Therefore more stringent filtering was necessary before a FlexX run. When docking with LIDAEUS it was quicker to dock a very large database, filter the top scoring hits, and manually examine compounds that failed a specific parameter. The workflow for each library examined is described in section 7.7.

7.6.1 Excluding compounds with undesirable functional groups

The physicochemical parameters used in the rules described in section 1.4 focuses on a few global parameters for each molecule. The functional groups within the molecule can inform

the prediction of metabolism, excretion and toxicity (Dearden, 2003). Databases, such as EDULISS 2, contain compounds with undesirable reactive groups; these compounds are often building blocks for chemical synthesis. Other fuctional groups have been implicated as causing toxicity or mutagenicity. The pharmaceutical industry keeps detailed records and monitors the literature for functionality that has caused problems in past drug development programs. The program DEREK (Deductive Estimation of Risk from Existing Knowledge) employed as part of EDULISS 2 is an expert system derived from the knowledge of human experts. As part of my CASE Studentship it was possible to filter virtual libraries and remove compounds with undesirable functionality. A proprietary algorithm was used to flag compounds green, amber or red depending on the presence of certain functional groups.

7.6.2 Predicting solubility

A low value of logP is indicative of a compound having reasonable aqueous solubility. Algorithms for predicting aqueous solubility from structure almost universally rely on a directly proportional relationship between logP and solubility (Jorgensen and Duffy, 2002; Delaney, 2005). The general solubility equation (described by equation 30) is a simple empirical relationship that predicts aqueous solubility from logP and melting point.

$$\log S_w = 0.5 - 0.01\ (T_m - 25)\ - \log P \qquad \textbf{Equation 30}$$

Where: S_w is the aqueous solubility in units of $mol.L^{-1}$, T_m the melting temperature in ^oC and logP the log of the octanol/water partition coefficient P.

Unfortunately, experimentally measured values for T_m and logP are rarely available. This significantly reduces the usefulness of this equation! It was not used in this study. In this study MlogP was used to guide compound selection. The general cut-off value for MlogP was set at 4.5. Compounds with MlogP $\leq$ 2 were usually soluble (~90 %). 2 $\leq$ MlogP $\leq$ 4.5 was less predictive (50 – 60 %).

7.7 The 4 screening collections

7.7.1 The Fragment library

The Maybridge Ro3 fragment library

The Maybrige Ro3 (www.maybridge.com/images/pdfs/ro3frag.pdf, November 2008) fragment library is a commercial fragment library of 1000 high purity ($\geq$ 95 %) compounds.

The compounds all satisfy the “rule of 3”, a refined version of Lipinski’s “rule of 5”. Satisfying the “rule of 3” should increase the probability of good ADMET properties (Lipinski *et al.*, 1997; Carr *et al.*, 2005). Compounds were selected to show good chemical diversity and to exclude functional groups that could lead to unsuitable reactivity or lack stability in aqueous solution. The fragments do however exhibit functionality that can be useful to link fragments to other chemical moieties during lead optimisation. The aqueous solubility of compounds was computed using the ALogGPS algorithm (Tetko and Tanchuk, 2002) was > 1 mM. The computed octanol:water partition ClogP (calculated logP) was provided for each compound using a proprietary algorithm (Chemomine Consultancy Limited).

It was not possible to purchase the whole Maybridge Ro3 library for reasons of cost. 96 compounds were selected from the 1000 compounds by Dr. Matt Nowicki (Edinburgh University) based on diversity. The structures of the 96 compounds are shown in Appendix 2. Maybridge supplied the 96 compounds on a 96-well plate in 1mg aliquots.

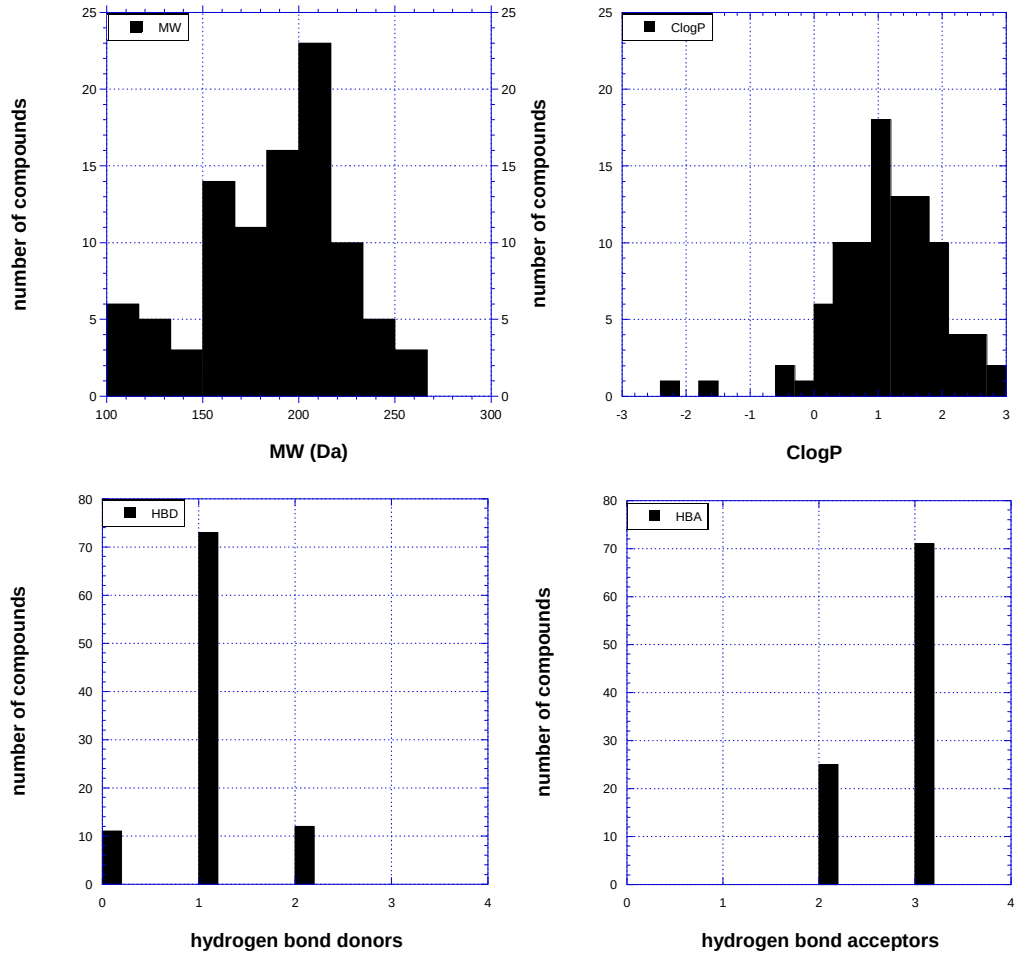


Figure 7-5 Distribution of calculated properties in the fragment library

Histograms showing the physicochemical properties for the fragment library. Properties were calculated by the supplier using commercial packages (Daylight, Chemical Information Systems Inc. and Chemomine Consultancy Limited).

The fragment library provided a very useful series of compounds to assess the thermal denaturation fluorescence assay (TDF).

Materials and Methods

The fragments were screened by TDF assay. Compounds were dissolved in 50 % double deionised water: 50 % DMSO mixture to a concentration of 50 mM in the well of the plate supplied by the manufacturer. 94 of the 96 compounds dissolved to 50 mM, two compounds dissolved after diluting with water to 25 mM. The plate was sealed with aluminium film and stored at -80 ^{o}C (Adhesive aluminium foil seals rated to -80 ^{o}C, Thermo scientific).

Screening sub-plates (ABgene PCR plates, rated to -80 ^{o}C, Thermo scientific) were made up with 100 mM bis-tris buffer at pH 7 with a fragment concentration of 625 μM (300 μl/well). The TDF assay was performed at a ligand concentration of 250 μM. The 625 μM plate provided the aliquots of 20 μl necessary for 3 repeats and 1 fluorescence control. The arrangement of the plate was designed to facilitate the use of an 8-tip multi-channel pipette. This reduced random error during pipetting into the main screening plate. For the same reason, solutions of protein and SYPRO orange were pre-diluted in bulk with buffer to ensure that pipetted volume was always ≥ 10 μL. The final well volume was 50 μL.

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		F41	F42	F43	F44	F45	F46	F47	F48	SD1_6	Neg_6	
C		F33	F34	F35	F36	F37	F38	F39	F40	SD1_5	Neg_5	
D		F25	F26	F27	F28	F29	F30	F31	F32	SD1_4	Neg_4	
E		F17	F18	F19	F20	F21	F22	F23	F24	SD1_3	Neg_3	
F		F9	F10	F11	F12	F13	F14	F15	F16	SD1_2	Neg_2	
G		F1	F2	F3	F4	F5	F6	F7	F8	SD1_1	Neg_1	
H												

Figure 7-6 Fragment plate for the TDF assay

48 compounds and controls were tested in parallel on a TDF plate. A similar plate array was set for compounds 49-96.

The TDF screening plate is shown in Figure 7-6. 48 fragments (250 μM), were tested in parallel on a single plate. The fragments are labelled F1 to F48 in Figure 7-6; fragments 49-96 were tested on a similar array. 6 positive controls (SD_1-6) and 6 negative controls (Neg_1-6) were included on each plate. The final solution contained 10 μM FKBP12, 250 μM Ligand/SD1$_{(+ \text{ control})}$; 30 mM Bis-Tris and 40 mM ammonium acetate at pH 7; 0.25 % DMSO, 0.25 % methanol and 10x SYPRO orange (SYPRO orange has a proprietary structure and is supplied as a 5000x solution in DMSO). The edge wells were kept empty. There was no evidence in this project that wells close to the edge gave less good results. Each plate was repeated in triplicate. A fluorescence control plate was also carried out. Conditions for the fluorescent control were similar to the fragment screening plates; protein solution was replaced with the appropriate buffer. Plates were incubated at room temperature for 15 minutes prior to the start of every assay.

7.7.2 The KB series

The KB series of 20 ligands were synthesised by Dr. Kevin Bailey (University of Manchester). The aim of the project was to design a series of ligands that would target both FKBP12 and human cyclophilin A (hcypA).

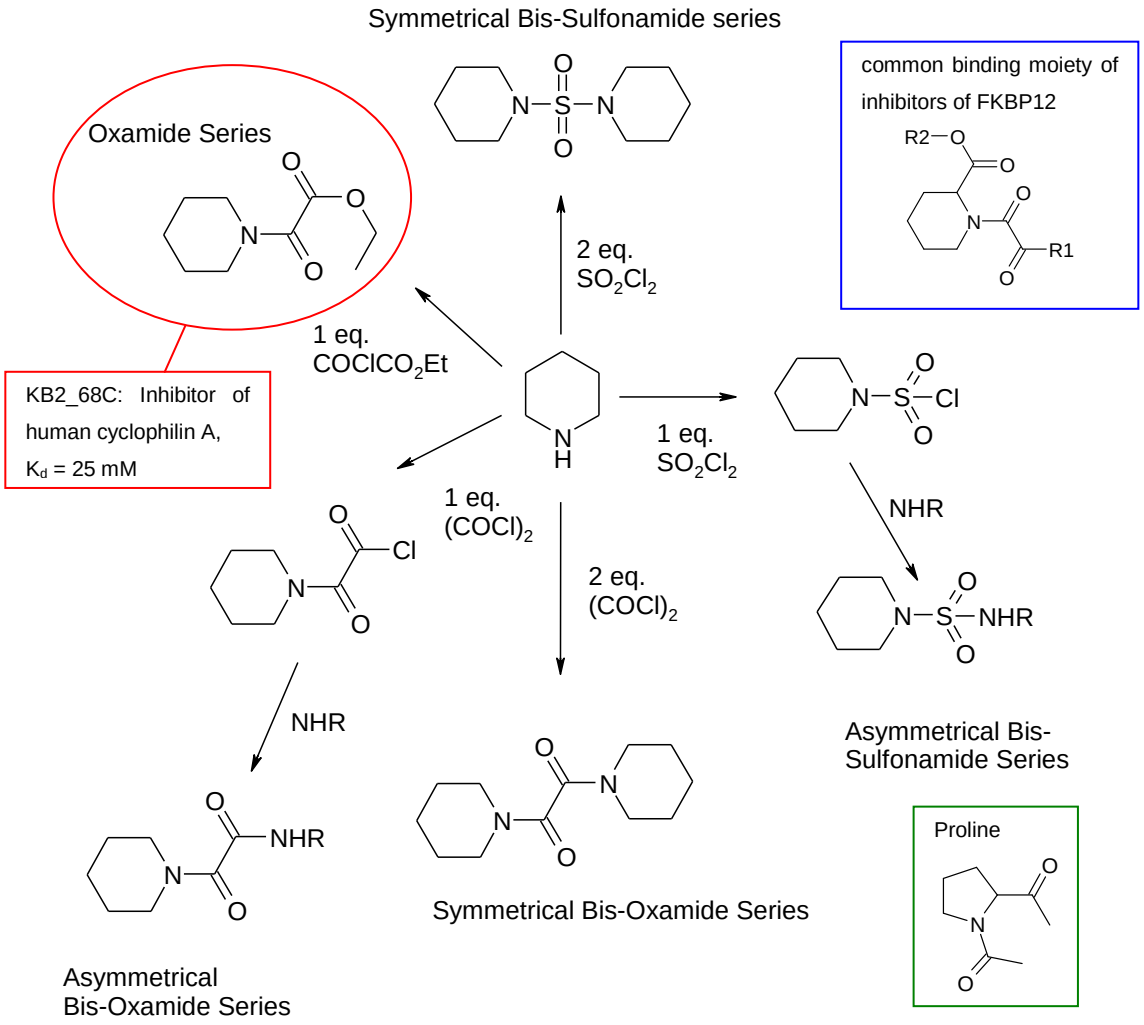


Figure 7-7 Synthesis of the KB series

The KB series were synthesised by Dr. Kevin Bailey from piperidine or an analogue of piperidine. Ethylpiperidine glycoxylate (KB2_68C) is circled in red, this compound has been found to inhibit cyclophilin A with a K_d = 25 mM (Kontopidis *et al.*, 2004). The blue inset box shows a binding moiety common to many FKBP12 inhibitors (Figure 1-10 Core binding motif of many FKBP12 inhibitors). The green inset shows proline. FKBP12 and hcypA both act as peptidyl-prolyl isomerases.

The starting point for the synthetic strategy was guided by the following observations from the literature. In an x-ray crystallographic and biochemical study by Kontopidis *et al*, ethylpiperidine glycoxylate inhibited hcypA with a K_d = 25 mM (KB2_86C in this study, the compound circled in red in Figure 7-7. (Kontopidis *et al.*, 2004)). hcypA is the binding

partner of the immunosuppressant drug cyclosporin (Handschumacher *et al.*, 1984). The role of cyclophilin in vivo is not fully understood. However, like FKBP12, it's function as a peptidyl-prolyl isomerase is thought to be important in processes where protein folding/unfolding is a rate-limiting step in cell biology and when the cell is under stress (Galat, 2003). Both proteins have been shown to bind to di-peptides containing proline or small molecule ligands with a motif that mimics proline (Dornan *et al.*, 2003).

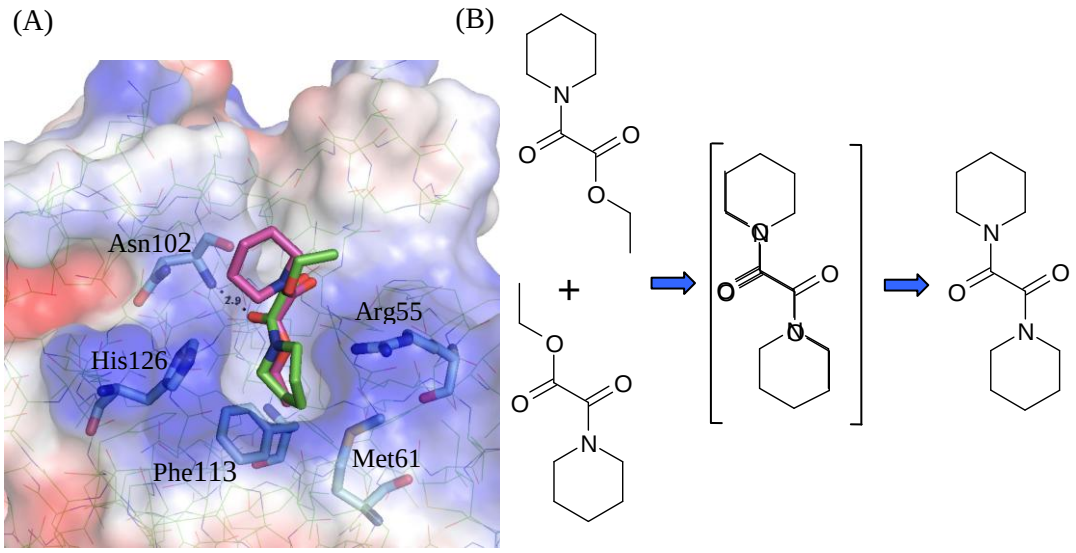


Figure 7-8 Ethylpiperidine glycoxylate in the active site of human cyclophilin A

(A) X-ray structure of the hCypA-KB2_68C complex (1W8M.pdb, Kontopidis *et al* (2004)). Two orientations of the ligand are shown: purple C-atoms 65 % occupancy, green C-atoms 35 % occupancy. An electrostatic surface was generated to illustrate the geometry of the active site; PyMol V1.02 (DeLano Scientific). The hydrogen bond between C=O and the main chain nitrogen of Asn102 is labelled with a black dashed line, 2.9 Å (B) The symmetrical bis-oxamide ligands were synthesised in an aim to increase the potency of KB2_68C, towards cyclophilin A, by simultaneously making the interactions of both orientations of KB2_68C seen in the active site.

The second observation was that the common binding moiety of many inhibitors of FKBP12 is a pipecolate ring with a diketo linker from the nitrogen on ring and an ester linker in the ortho position (blue inset box in Figure 7-7); this is structurally similar to KB2_68C (Dornan *et al.*, 2003; Wang and Etzkorn, 2006). The binding mode of this moiety for FKBP12 is discussed in detail in section 1.11.6. The aim of the KB series of compounds was to test if it were possible to design inhibitors to inhibit both hcypA and FKBP12. Examining the structures of known actives of FKBP12 would suggest that modifying KB2_68C by making substitution at the *ortho*/*meta* positions of the piperidine ring might lead to a compound that would target FKBP12. Substitutions at the *ortho/meta* positions might however have positive or negative implications for the activity of KB2_68C analogues against hcypA. Ligands of

FKBP12 are generally larger than those of hcypA; a direct consequence of the dimensions of the active site. hcypA has a narrow trench-like binding groove (Figure 7-8) compared to that of FKBP12 which is a wide hydrophobic pocket. The crystal structure of KB2_68C in complex with hcypA is particularly interesting; the ligand is seen in two positions in the active site (Figure 7-8 (A)). In both poses the ligand makes a hydrogen bond with the main chain nitrogen of Asn102; this is seen in all peptide-hcypA structures (Ke *et al.*, 1991; Kallen *et al.*, 1998). The occupancy of the ligand shown with purple C-atoms was 65 %, the ligand shown with green C-atoms, 35 %. This is interesting as the purple C-atom ligand is the least buried of the ligands. This crystal structure suggested that a symmetrical bis-oxamide might be more active than KB2_68C. Figure 7-8 (B) illustrates how a symmetrical bis-oxamide might simultaneously make the interactions seen in the two poses of KB2_68C observed in the active site. It is worth noting that the piperidine ring of the green KB2_68C fits into a fairly small pocket adjacent to Phe113 and modifications to the piperidine ring might lead to changes to the amino acid side-chain conformations lining the pocket (Arg55, Met61, Phe113, and His126). The aim was to increase the potential for non-covalent interactions between ligand and protein and improve ligand affinity.

An additional aim was to design soluble compounds. The law of mass action dictates that ligand solubility is of paramount importance when characterising the binding of low affinity ligands. The KB series of molecules range in mass from 185.2 to 292.4 Da.

Synthetic strategy

Figure 7-7 shows the synthetic strategy which is based on the reaction of piperidine with various acid chlorides. These reactions lead directly to symmetrical bis-sulfonamides, symmetrical bis-oxamides, an asymmetric oxamide or an acid chloride that was converted to an asymmetric bis-oxamide or sulphonamide by reaction with an imine. Replacing the piperidine with a piperidine substituted at one or more positions on the ring, morpholine, 1,4-oxathiane or pyrrolidine expanded each series. The oxamide, bis-oxamide and bis-sulphonamide series are illustrated in Figure 7-9, Figure 7-10, Figure 7-11 respectively.

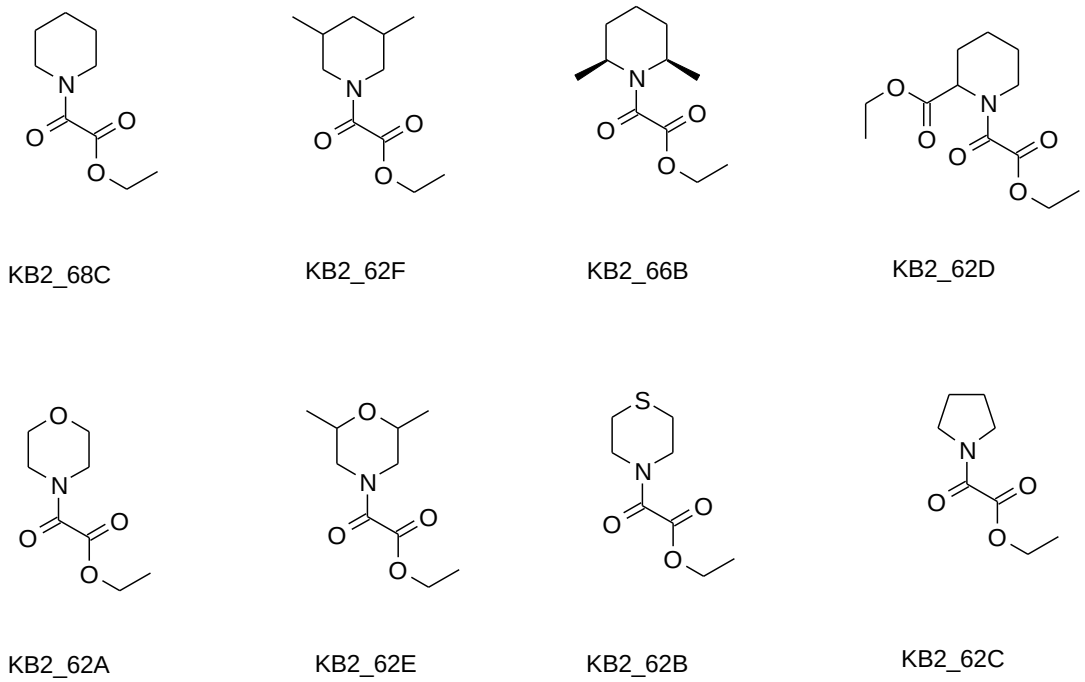


Figure 7-9 KB oxamide series

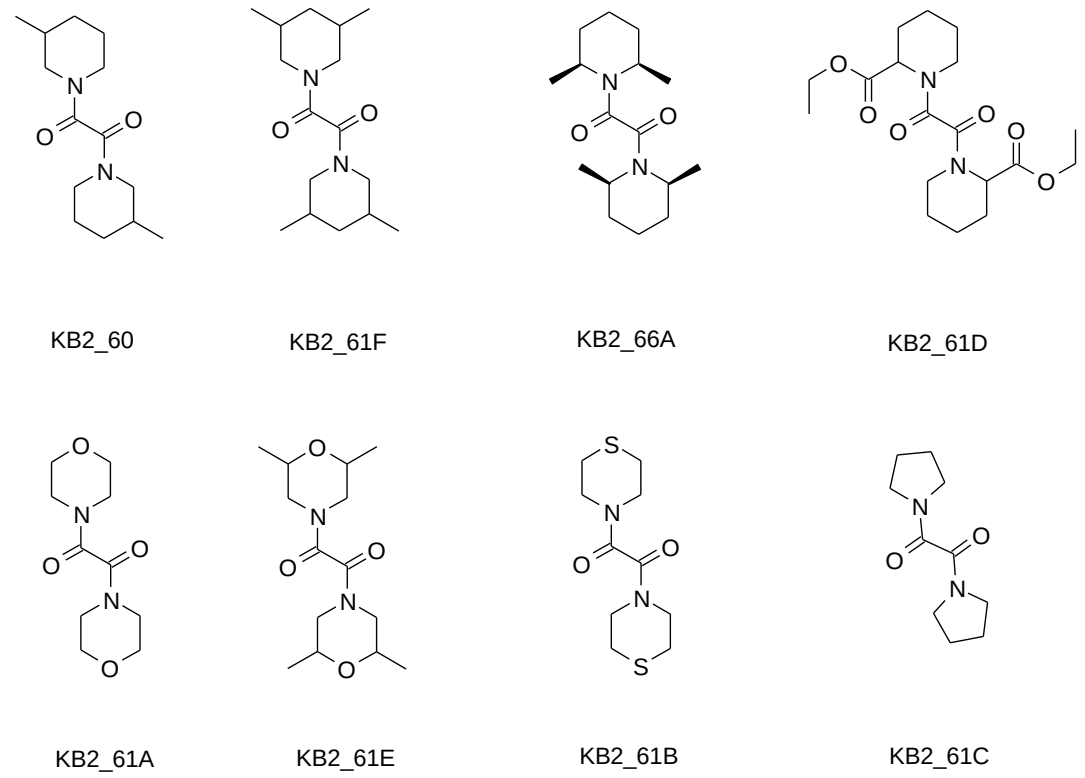


Figure 7-10 KB bis-oxamides

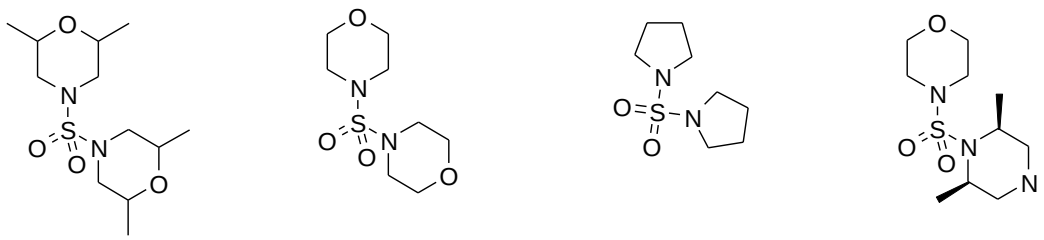


Figure 7-11 KB bis-sulphonamides

Materials and Methods

Initial Screen -TDF

The initial screen for the KB series was by TDF assay. Stock solutions were made by dissolving the compounds in HPLC grade methanol (BDH) to 100 mM. Compound KB2_61B was the least soluble compound and a stock solution of 25 mM was made. Solutions were divided into 100 µl aliquots and stored at -80 ^{o}C in cryogenic vials (Nunc). 96-well plates were made up of the compounds in 100 mM bis-tris buffer at pH 7 at a concentration of 625 µM following similar protocols to those described for the fragment screen (300 µl/well). The plates were sealed with aluminium foil and stored at -80 ^{o}C.

The 20 compounds were tested in parallel on a single plate. 3 positive controls (SD1) and 3 negative controls were included on each plate. The final solution contained 10 µM FKBP12, 250 µM Ligand/SD1 (+ control); 30 mM Bis-Tris and 40 mM ammonium acetate at pH 7; 0.25 % methanol and 10x SYPRO orange (SYPRO orange has a proprietary structure and is supplied as a 5000x solution in DMSO). The experiment was repeated in triplicate in separate experiments. A fluorescence control was also included for each compound on the plate. Plates were incubated at room temperature for 15 minutes prior to the start of the assay.

Confirming the specificity of KB2_61D for the active site of FKBP12 - PPIase

A peptidyl-prolyl isomerase assay (PPIase) was performed to determine the specificity and the K_d of KB2_61D (most active member of the series) by competitive inhibition of the isomerisation of the substrate Succ-Ala-Leu-Pro-Phe-*p*-nitroaniline by FKBP12. A titration was carried out using compound KB2_61D at concentrations between 0 and 500 µM using the protocol described in section 6.3.1. The apparent equilibrium dissociation constant was

estimated from the best fit of equation 20 to the rate of turn-over of the substrate, corrected for the background turnover at 4 ^{o}C between 0.5 and 1.5 s after addition of the substrate, plotted against the concentration of the inhibitor as described in section 6.4.1. The equilibrium dissociation constant, K_d, for each inhibitor, was determined by correcting the K_i for competition with the substrate using equation 21.

Thermodynamic characterisation of the FKBP12-KB2_61D complex

An isothermal titration calorimetry (ITC) was carried out to characterise the thermodynamic parameters for the FKBP12-KB2_61D complex (ITC is described in section 6.3.2). ITC experiments were carried out on a MicroCal auto-VPITC instrument (GE Healthcare). Reagents were dissolved in 50 mM sodium phosphate, 50 mM sodium chloride at pH 7; 0.25 % methanol. 300 μM KB2_61D (syringe) was injected into 22 μM FKBP12 (reaction cell). The experimental protocol was 28 x 10 μl injections (0.5 μl 1st injection), 20 s injection duration, 180 s injection delay, syringe spin 240 rpm, reference power 5 μcals^{-1}, 25 ^{o}C. The experiment was repeated in triplicate including standard controls to account for the heat of dilution of the ligand and protein and the heat of mixing. Data analysis was as described in section 6.3.2).

Determining the K_d of the FKBP12-KB2_61D by ESI-MS

An ESI-MS titration was performed with 10 μM FKBP12 and 0 to 300 μM KB2_61D in 10 mM ammonium acetate at pH 6.8 on the LCQ^{TM}DECA (ThermoQuest) instrument using the protocol described in section 3.5.6.

Clustering the KB series by chemical similarity

The KB series was clustered by chemical similarity; the resulting dendrogram was used as a tool to examine the activity data provided by the TDF assay. Fingerprints were calculated from 2D representations of the molecules using the program Accord for excel (Accelrys). A pair-wise comparison was made of the bit strings for each molecule using a Tanimoto coefficient (Haranczyk and Holliday, 2008). The resulting symmetrical similarity matrix (20x20 for 20 compounds) was used as the input for a UPGMA clustering algorithm (MVSP V3.13, Kovach Computing Services).

Effect of KB2_61D on two strains of *C.elegans*

Sexually mature wild type (N2) and FKB3 triple mutant (TP83) *C.elegans* were exposed to 10 to 500 μM KB2_61D for 6 days; worm growth was compared to negative controls. KB2-KB2_61D was dissolved in ethanol; the final ethanol concentration was less than 3 % (v/v). These experiments were carried out by Dr. Victoria Butler (Faculty of Veterinary Medicine, University of Glasgow).

7.7.3 The SL series

Figure 7-12 illustrates the virtual screening strategy used for selection of the SL series.

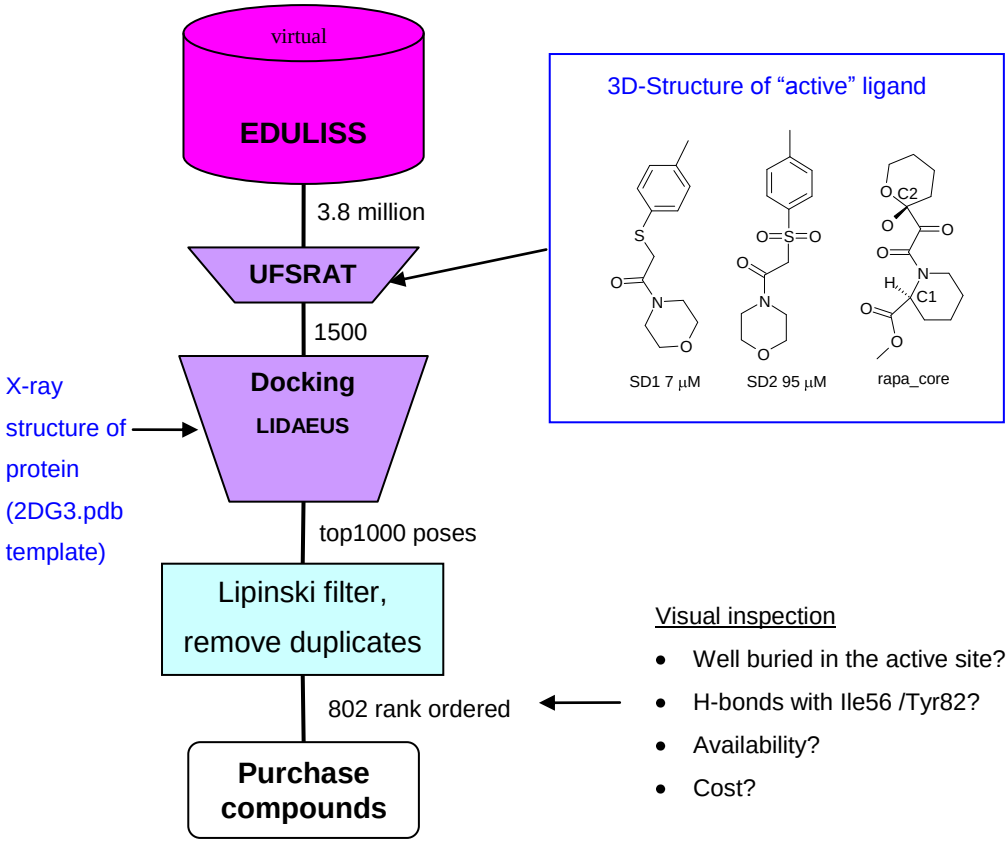


Figure 7-12 Selecting the SL series by virtual screening

The SL series of compounds were selected from the EDULISS 2 database using two virtual screening approaches in series; an algorithm that calculated 3D similarity coefficients for test compounds compared to a known active (UFSRAT) and a protein-ligand docking algorithm (LIDAEUS). Compounds were filtered for drug-likeness using Lipinski's criteria. This strategy reduced the compound set to 0.02 % of the original EDULISS 2 database. Docking poses were manually inspected for potential interactions with key residues in the active site.

The selection of the SL series was based on the premise that structurally similar compounds are likely to exhibit similar physiological effects (Crum Brown and Fraser, 1869; Bender and Glen, 2004). The choice of query structure for UFSRAT was based on activity and size. Poor ligand solubility is often a limiting factor in estimating affinity by biophysical methods for ligands with K_d > 10 μM. This has been a problem throughout this study. The activity of a compound relative to size is worthy of particular consideration. The aim was to pick a structure with high ligand efficiency. It was felt that use of a ligand with pendant groups that

made few non-covalent interactions with FKBP12 would lead to poorer enrichment by UFSRAT.

Three structures were used to query the EDULISS 2 data base, SD1, SD2 and the core binding motif of rapamycin. SD1 and SD2 are known ligands of FKBP12 that bind specifically to the active site (Stebbins *et al.*, 2007). The nanomolar affinity immunosuppressant ligands of FKBP12; rapamycin and FK506 are macrocyclic lactones; ~50 % of the ligand binds not to FKBP12, but to a second protein (mTor and calcineurin respectively). It is the inhibition of this partner protein that leads to their immunosuppressive action. Our aim was to inhibit FKBP12 without immunosuppression. For these reasons rapamycin and FK506 are unsuitable query structures for a similarity search. However, both share the same FKBP12 binding motif (Figure 1-7). This motif occupies the majority of the active site making 4 hydrogen bonds with key residues. The first approach was to select the fragment of rapamycin that contained atoms within 3.5 Å of FKBP12 (PyMol V1.02, Delano Scientific). The rationale being to create a "shape" made up of atoms of the correct type (atoms with the potential to make non-covalent interactions) and spatial distribution matching hot spots in the active site. This is analogous to the generation of site-points for a docking experiment. This approach was refined when docking experiments showed that the compounds selected by UFSRAT were generally too massive to dock deeply into the active site of FKBP12. During manual inspection of the docked ligands, very few made hydrogen bonding interactions with residues identified as being hot spots in the active site (hydroxyl of Tyr82 or main chain N of Ile56). The rapamycin query structure was pruned to include only the most buried atoms (rapa_core, Figure 7-12). Pruning retained atoms making hydrogen bonds with FKBP12.

The 3D coordinates of SD1, SD2 and those in EDULISS 2 were generated using the program CONCORD (TRIPOS). The conformation of the query structure is clearly of considerable importance. One might argue that this should be that of the ligand in complex with the protein. Alternatively, one could argue that the 3D structures from EDULISS 2 generated by CONCORD are minimised low energy conformations *in vacuo* and that it is this conformation of the ligand that should be used. The former approach should favour the retrieval of more rigid scaffolds that closely match the shape of the active site. The latter approach has been shown to give the query compound the highest similarity score (Shave and Blackburn, unpublished results). Care was taken to make sure that the two chiral centers in rapa_core matched those in rapamycin and FK506. These atoms are labelled C1 and C2 in

Figure 7-12. This was done by taking the conformation of rapamycin from an x-ray structure of a complex with FKBP12 (2DG3.pdb). There are no x-ray structures available for SD1 or SD2 in complex with rapamycin and a low energy conformer was generated using the program CONCORD (TRIPOS).

The top 500 hits from UFSRAT were saved for each query structure (1500 total) and docked using the programs LIDAEUS. The generation of the site-points defining the active site of FKBP12 is described in section 7.2.3 (grid of spacing of 0.09 Å for the site-points). The docking resolution parameter was set at 0.04 Å. This parameter setting was arrived at by re-docking a minimised conformation of rapa_core at 0.02, 0.04 and 0.06 Å. Setting the resolution to less than 0.04 Å meant that the program failed to rank the pose most similar (low RMSD) to that seen in the crystal structure (2DG3.pdb) as the highest in a series of solutions. The highest ranked pose of a complete structure of rapamycin has an RMSD of 1.2 Å when compared to the crystal structure. Setting a more relaxed docking resolution (larger) is a fast method of accounting for a small amount of conformational flexibility in the ligand. Ligands matching the shape and atom types of the site-points were ranked according to the enthalpy of the interaction and buriedness. The top 500 poses were saved from each query. These poses were analysed for duplicates and the highest scoring pose saved, when more than one solution were available. The set of unique compounds was filtered for drug-likeness using Lipinski's rule of 5 criteria (Lipinski *et al.*, 1997). The compounds were also docked using the program FlexX. These series of in silico experiments reduced the 3.8 million unique compounds in EDULISS 2 to 802 compounds (0.02 % of the database). The poses were then visually examined for buriedness and putative interactions with important residues in the active site. Ligands were only selected if they occupied the base of the pocket and made at least one hydrogen bond with Tyr82 (hydroxyl) or Ile56 (main chain N). Compounds that met these criteria were grouped by supplier and their availability and cost determined. When a compound was not available the search engines in PubChem http://pubchem.ncbi.nlm.nih.gov, December 2008, Chembase (Seiler *et al.*, 2008) and ZINC (Irwin and Shoichet, 2005) were used to source an alternative supplier. 10 compounds were purchased.

To conclude, the aim in the selection of the SL series was to use shape similarity to select a similar, but "not too similar compound". UFSRAT has a greater potential to scaffold hop than the use of a 2D binary fingerprint that encodes specific structural elements in the query structure (Boehm *et al.*, 2004; Ballester and Richards, 2007).

Materials and Methods

Initial Screen -TDF

The SL series were tested using the thermal denaturation fluorescence assay, as described in sections 6.3.4 and 7.8.1. The compounds were supplied as a solid, they were dissolved to a concentration of 30 mM in methanol (SL1, SL2, SL3, SL7 and SL9) or DMSO (SL5, SL6, SL8), and stored at -80 ^{o}C. 8 of the 10 compounds purchased dissolved. No suitable solvents were found for SL4 or SL10. Compounds were screened at 10 μM FKBP12 and 250 μM screening compound. Compound SD1 250 μM was used as the positive control; K_d = 6.8 μM. All wells contained equal concentrations of DMSO (0.8 %) and methanol (1 %) The assay was repeated in triplicate on separate plates.

ESI-MS

SL1, SL2, SL3, SL7, SL9 were tested at 100 μM ligand, 10 μM FKBP12 in 10 % methanol (v/v), 10 mM ammonium acetate, pH 6.8 on the LCQ^{TM}DECA (ThermoQuest) instrument using the protocol described in section 3.5.1. It was not possible to test SL5, SL6 and SL8 by ESI-MS due to the presence of DMSO.

Clustering the SL series by chemical similarity

The KB series was clustered by chemical similarity; the resulting dendrogram was used as a tool to examine the activity data provided by the TDF assay.

7.7.4 The Specs library

Compound selection

In this study a subset of 15000 compounds from the Specs catalogue (www.specs.net) were screened against FKBP12. 2D representations of 25 000 compounds from the supplier were processed using proprietary software when on placement with my CASE Studentship sponsor Organon (Part of the Schering-Plough Corporation, www.schering-plough.com, Jan 2009). SDF files of 25000 compounds were computationally cleaned; this process involved the removal of metal ions, fragments and the assignment of charge appropriate to pH 7.4. Compounds were compared to remove duplicates and than assessed for drug likeness. The term drug-likeness refers to passing Lipinski's Ro5 (Lipinski *et al.*, 1997) and absence of undesirable functional groups. 14894 compounds passed the most stringent set of criteria; these are referred to as the Specs green subset. Figure 7-13 shows the distribution of MW, ClogP, number of hydrogen bond donors and hydrogen bond acceptors. 3D SDF files were created using the program CONCORD (TRIPOS).

The 14894 compounds were docked using the programs LIDAEUS and FlexX using the crystal structure of FKBP12 in complex with rapamycin as a template (2DG3.pdb). UFSRAT was used to select compounds most similar to 3 known actives as described in section 7.7.3. The Bayesian algorithm of Dr Nicos Angelopoulos (Angelopoulos *et al.*, 2009) was used to assign a probability of activity against FKBP12 for each of the compounds (The San Diego Data Set was used as the training set; http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=608, January 2009). The 40 top scoring hits were selected from each program and those that had a good score in LIDAEUS and FlexX (ranked by the arithmetic addition of scores from each program). 10 of the compounds were selected by more than one program. Compounds were also selected from the Organon compound collection (all purchasable compounds) that contained fragments that were known to bind to FKBP12 (crystallographic data, Appendix 3). This was carried out by means of a BCI fingerprint.

Materials and Methods

In total 162 compounds were tested using the thermal denaturation assay, as described in section 6.3.4. The compounds were supplied at 10 mM in DMSO and screened with 15 μM FKBP12 and 250 μM test compound. Compound 250 μM KB2_61D was used as the positive

control; K_d = 6.6 μM. 40 compounds were tested on each plate with 5 repeats for positive and negative controls. To date the whole compound collection has been tested twice.

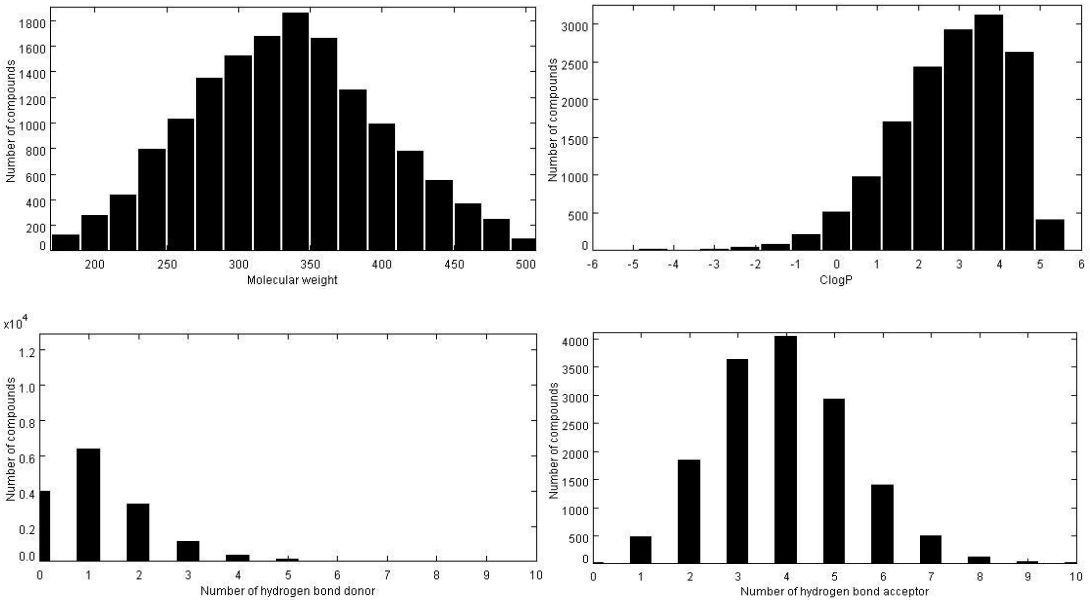


Figure 7-13 The Specs green subset

Distribution of MW, ClogP, number of hydrogen bond donors and hydrogen bond acceptors for the Specs green subset. Histograms were created using the program Topcat (www.star.bris.ac.uk/~mbt/topcat/; January 2009).

7.8 Results

7.8.1 The Fragment library

Figure 7-14 illustrates the results of the fragment screen by TDF assay.

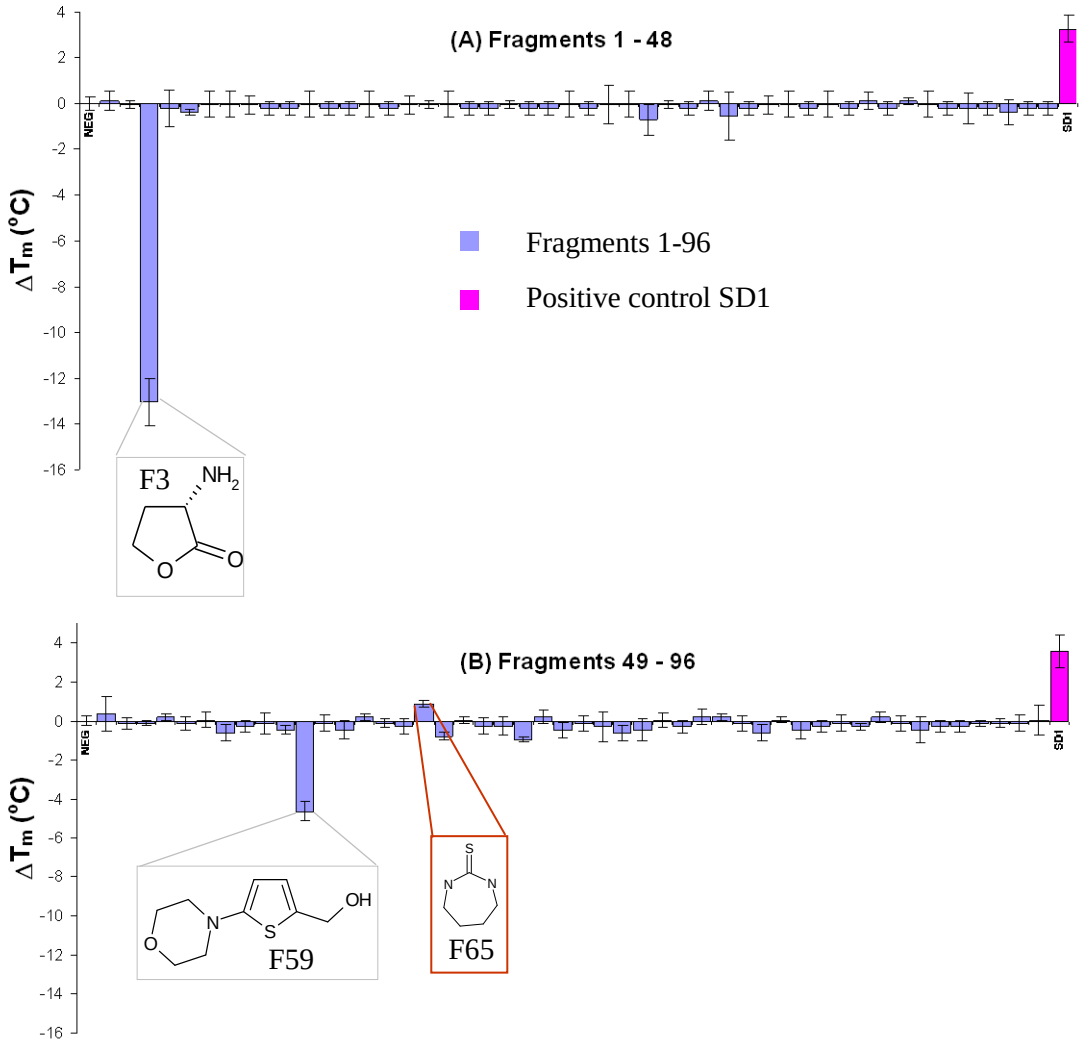


Figure 7-14 TDF screen of the fragment library

TDF screen of the fragment library. The y-axis shows the shift in mid-point melting temperature (ΔT_m °C). 10 µM FKBP12; 250 µM fragment; 10x SYPRO orange, 0.25 % DMSO, 0.25 % methanol, 30 mM ammonium acetate; 40 mM Bis-Tris, pH 7. Positive control contained 250 µM SD1 in place of a fragment. The negative control contained no ligand. (A) Fragments 1-48. (B) Fragments 49-96. Two compounds reduced the mid-point melting temperature by a significant amount: F3 and F59, these are highlighted in the figure. F65 gave a small positive melting point shift; this compound is boxed in red.

One compound gave a $+\Delta T_m > 0.5$ °C, 1,3-diazepane-2-thione (Fragment F65, $\Delta T_m = 1.3$ °C). This would equate to a K_d of between 40 and 70 µM at 25 °C (extrapolating

to 25 ^{o}C for ΔH_L using equation 24, ΔC_p was not included). Two simple cyclic compounds, cyclohexanone and cycloheptanone, were identified in Burkhard's study of ligands binding to FKBP12 with K_d of 50 μM and 30 μM respectively (Burkhard, PhD Thesis, Basel University, 1995). Burkhard's study measured binding by the quenching of Trp59 fluorescence as function of ligand concentration (Park *et al.*, 1992). In a study by Kontopidis, cyclooctanone was crystallised in complex with FKBP12 (Kontopidis, PhD Thesis, Edinburgh University, 1999). Cyclootanone occupied a position close to Trp59 in the active site in a similar location and orientation to DMSO (1d7h.pdb, (Burkhard *et al.*, 2000)).

Two fragments: F3 ((S)-(-)-alpha-amino-gamma-butyrolactone hydriodide) and F59 ((5-morpholinothien-2-yl)methanol) gave negative ΔT_m of -13.1 ^{o}C and -4.6 ^{o}C respectively. It is interesting to note, fragment F3 also reduced the mid-point melting temperature of pyruvate kinase from *Leishmania mexicana* (ΔT_m ~ -5 ^{o}C) and reduced the activity of the protein in an activity assay. The effect of the compound was also tested in a crystallisation trial for pyruvate kinase where crystallisation conditions for the protein were well established. On addition of F3 there was rapid production of an amorphous precipitate (unpublished results; Dr. Hugh Morgan, Edinburgh University). For a small molecule F3 has good potential for making hydrogen bonds; protein destabilising effects may be explained by the compound acting as a chemical denaturant.

Trends in activity were investigated by examining relationships between ΔT_m and descriptor values. ΔT_m was plotted against MW, ClogP, number of hydrogen bond acceptors, number of hydrogen bond donors and polar surface area (Chemical Information Systems). No trends in activity were observed for any of the descriptors (results not presented).

Figure 7-14 shows the fragments that gave a small consistent positive ΔT_m; (these results are not statistically significant). The three compounds highlighted in grey show some structural similarity to known inhibitors of FKBP12. F62, F79 and F80 all contain a heteroaromatic ring linked by a sulfonyl group to a heterocycle. It is interesting to note that this substructure is seen to be the most deeply buried fragment in the ligand in complex with FKBP12 in the crystal structure 1J4I.pdb, (Sun *et al.*, 2003). Figure 7-16 shows the orientation of 308 in the active site of FKBP12. Appendix 3 shows the structures of ligands crystallised in complex with FKBP12.

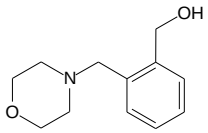
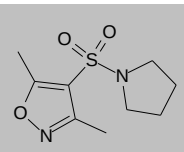
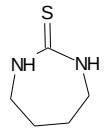
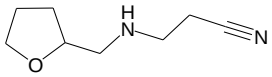
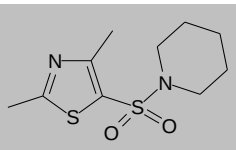
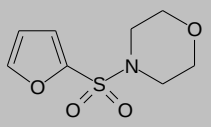
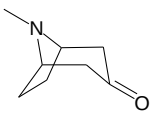
Fragment	*Structure*	ΔT (^{o}C)
F52		0.22 ± 0.17
F62		0.22 ± 0.17
F65		**1.29 ± 0.53**
F71		0.22 ± 0.17
F79		0.22 ± 0.39
F80		0.22 ± 0.39
F88		0.22 ± 0.17

Figure 7-15 Positive ΔT_m in the fragment screen with FKBP12

The majority of the compounds in the TDF screen gave a small – ΔT_m with FKBP12. The compounds shown above gave very small + ΔT_m. A ΔT_m = 0.2 ^{o}C would equate to a K_d apparent > 200 μM at the melting temperature; theses results are not statistically significant and are only included for the record. Fragment F65, ΔT_m = 1.29 ^{o}C, is the most interesting compound in the set, this equates to a K_d apparent of ~160 μM at T_m. This would extrapolate to between 40 – 70 μM at 25 ^{o}C (extrapolating to 25 ^{o}C for ΔH_L using equation 24, ΔC_p was not included).

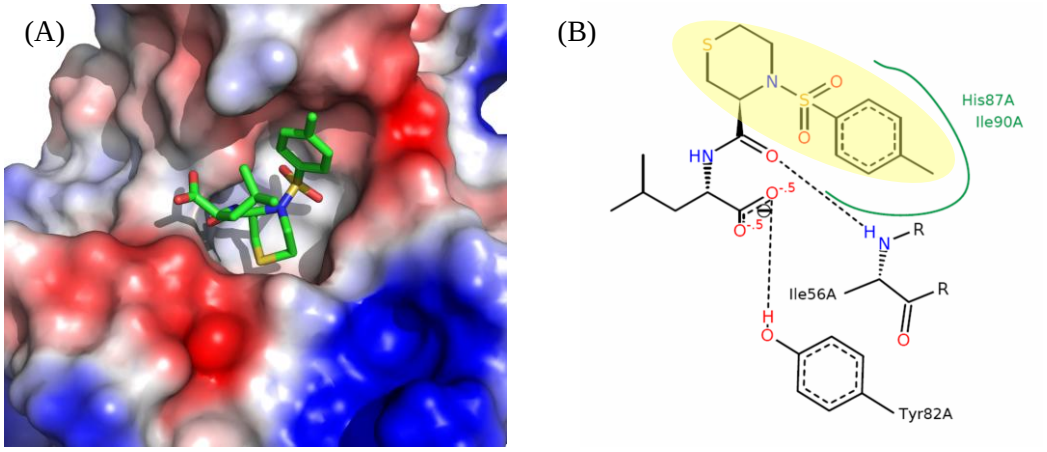


Figure 7-16 Compound 308 in complex with FKBP12.

Representation of the x-ray structure of compound 308 in complex with FKBP12 (1J4I.pdb, (Sun *et al.*, 2003)). (B) 2D schematic representation of the non-covalent interactions between the protein and ligand from the x-ray structure using the Poseview server: http://poseview.zbh.uni-hamburg.de/~poseview/poseview.php (Stierand *et al.*, 2006). Hydrogen bonds are represented as dashed lines and hydrophobic interactions as green arcs. The shaded ellipsoid highlights the moiety in compound 308 that is structurally similar to the compounds in the fragment screen that showed a positive shift in mid-point melting temperature (F62, F79 and F80).

7.8.2 **The KB series**

The results of the TDF assay on the KB series are shown in Figure 7-18. Results are grouped by chemical similarity. The most active compounds were KB2_61D and KB2_62D; these compounds gave a shift in mid-point melting temperature, $\Delta T_m > 1.5$ °C. Table 7-1 shows ΔT_m and K_d estimated at the melting temperature, T_m, (calculated with using equation 23, section 6.3.4). K_d were extrapolated to 25 °C using 3 putative values of ΔH_L (equation 24).

Ligand	ΔT_m (°C)	K_d (µM) (T_m)	K_d (µM) ΔH^a = -2.0 kcal·mol^{-1} (25 °C)	K_d (µM) ΔH^b = -5.0 kcal·mol^{-1} (25 °C)	K_d (µM) ΔH^c = -7.0 kcal·mol^{-1} (25 °C)
KB2_61D	6.8 ± 0.3	22.2	*16*	*10*	*7*
KB2_62D	1.8 ± 0.3	131	*99*	*65*	*49*

Table 7-1 The most active compound cluster in the KB series (FKBP12)

The apparent equilibrium dissociation constant are shown at the melting temperature and at 3 putative values of ΔH_L.

Compound KB2_62D is a sub-structure of compound KB2_61D. Figure 7-17 is a plot of –dF/dT against temperature for KB2_61D and KB2_62D plus controls.

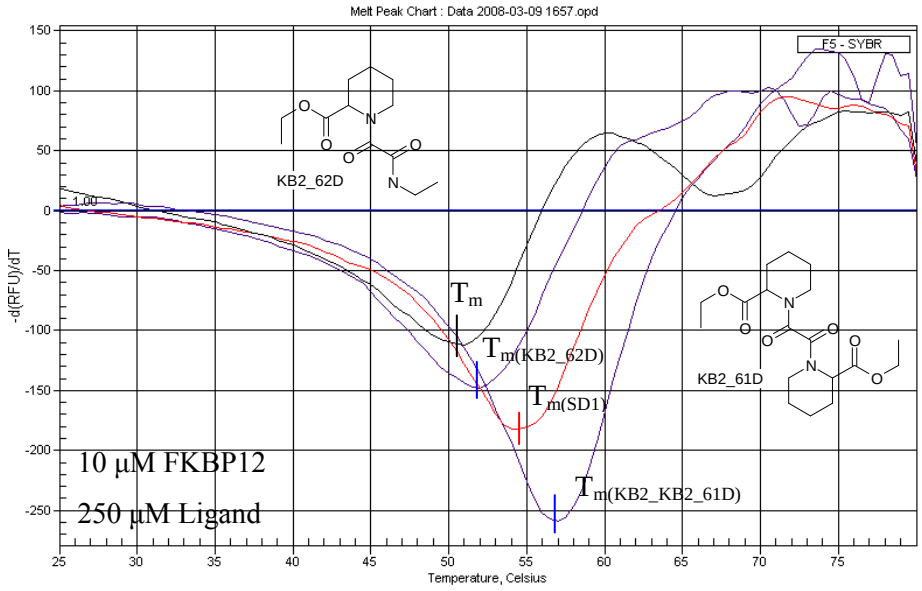


Figure 7-17 TDF data for compounds KB2_61D and KB2_ 62D and FKBP12

T_m is the minimum value of –dF/dT in the plot of –dF/dT against temperature. 10 µM FKBP12, 250 µM ligand, KB2_62D (purple), SD1 (red) and KB2_61D (blue). There is no ligand present for the black trace labelled T_m; 10x SYPRO; 0.25% methanol; 50 mM ammonium acetate, pH 7.

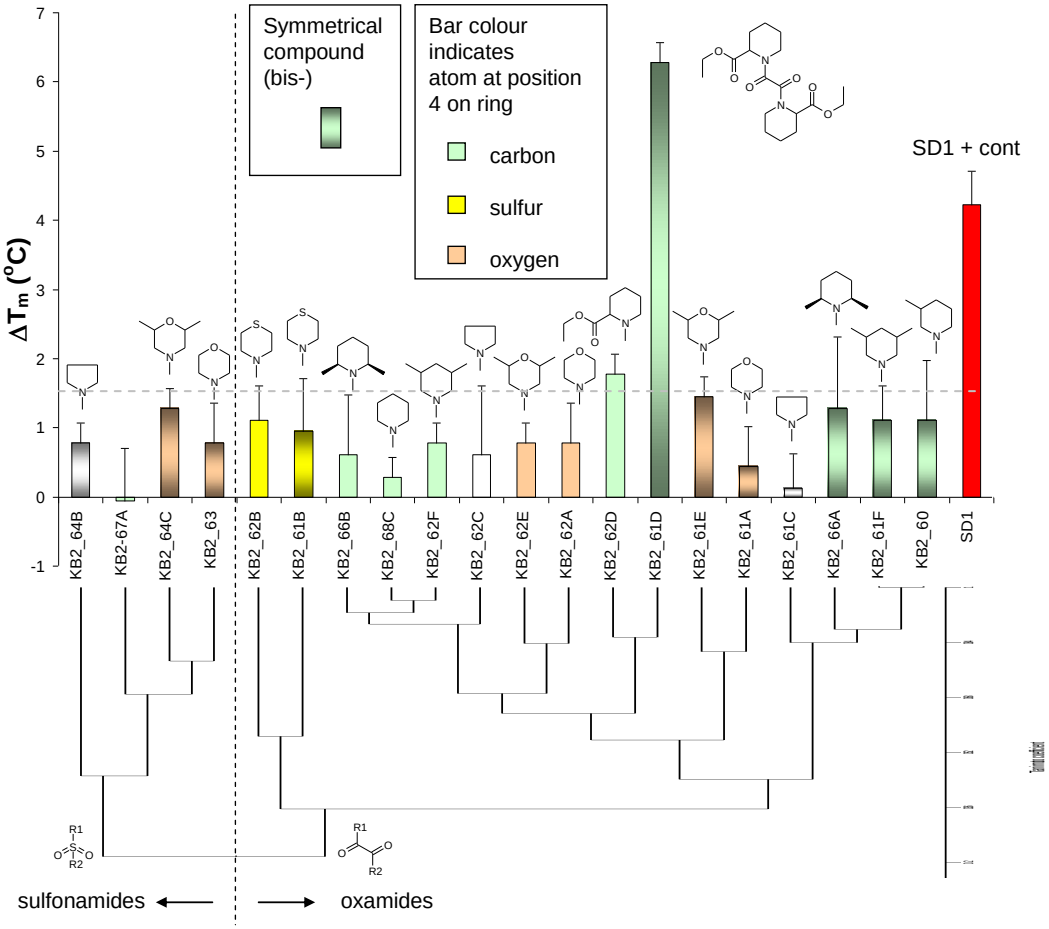


Figure 7-18 TDF assay results for the KB series of compounds

Compound KB2_61D stabilises FKBP12 to a greater degree than the 6.8 µM positive control SD1. 10 µM FKBP12, 250 µM ligand/SD1; 10x SYPRO orange, 0.25 % methanol, 30 mM ammonium acetate; 40 mM Bis-Tris, pH 7. Standard deviation is shown for n = 3 (three separate experiments). The structures of the heterocyclic ring are shown to illustrate the effect of changes to the ring (colour coded by atom type at position 4). The shaded bars represent symmetrical molecules with two identical rings. The cluster dendrogram is shown to illustrate similar compounds. The algorithm makes a primary classification by distinguishing between a sulphonamide or oxamide linker. See Figure 7-9, Figure 7-10 and Figure 7-11 for full ligand structures.

Structure activity relationships

In contrast to the screening of the fragment library, which provided only one compound with a ΔT_m of ~1 °C, the KB series gave 9 hits where ΔT_m was ≥ 1 °C. The most active cluster of compounds has the substituent -COOEt in the ortho position on the heterocyclic ring. KB2_61D and KB2_62D are the closest analogues of the core binding motif of known FKBP12 inhibitors. The modelling of compound KB2_61D into the active site of FKBP12 was discussed in some detail in section 4.4.4; Figure 7-19 shows a 2D schematic diagram of predicted interactions between ligand and protein. These compounds have the potential to make 2 hydrogen bonds, one between a carbonyl in the di-keto linker and Tyr82 and another

with the ester carbonyl on the closest ring and the main chain nitrogen of Ile56. In this binding mode the pipecolate moiety mimics proline in preferred substrates of FKBP12.

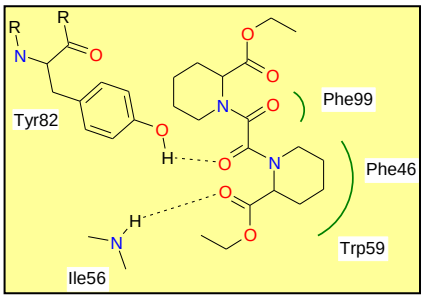


Figure 7-19 Model of KB2_61D in the active site of FKBP12

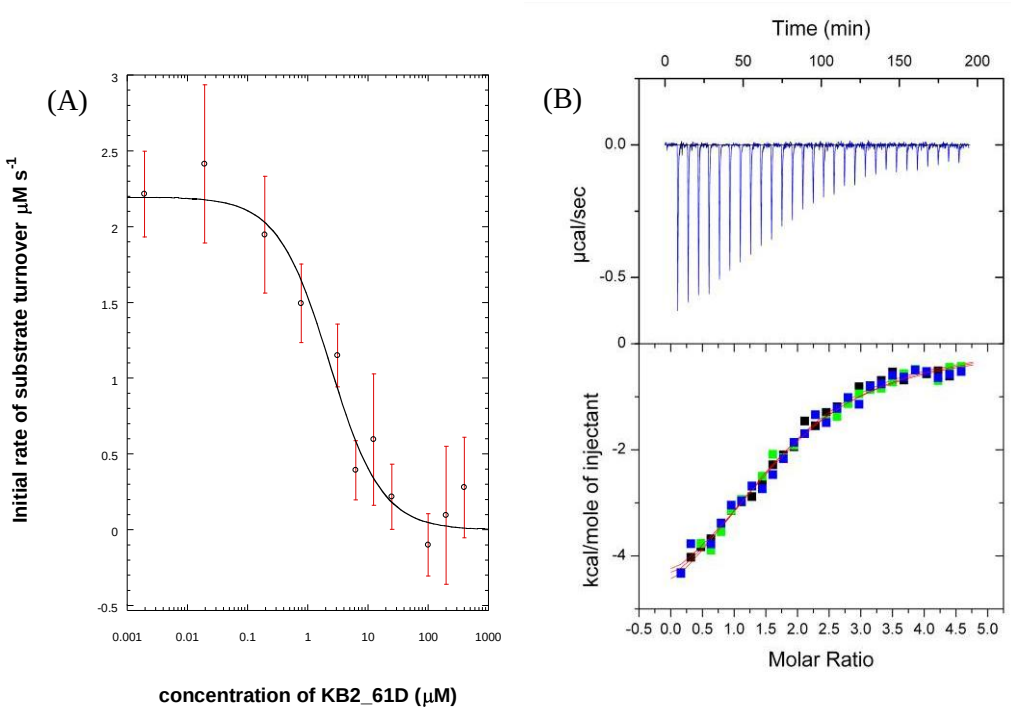
2D schematic diagram of electrostatic interactions predicted by the model of KB2_61D in the active site of FKBP12. Protein-ligand electrostatic interactions are represented by dashed lines and hydrophobic interactions by solid green arcs. Not all hydrogen atoms are shown for clarity. FlexX V3.0.2 (Rarey *et al.*, 1996). The docking is described in section 4.4.4.

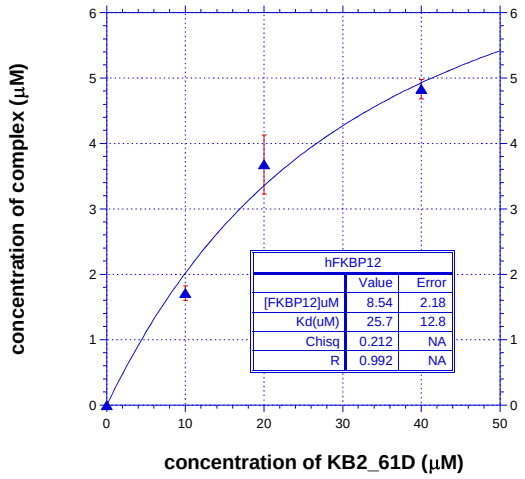
The symmetrical compounds containing two heterocycles are more active than those with one heterocyclic ring. Changes to the atom in position 4 on the heterocycle from C to S or O does not seem to have a significant effect on activity. The general trend in activity with substitution in the ortho/meta position on the heterocyclic ring is: –COOEt > CH_3 > H. There are examples in the series with a methyl group in the ortho positions and in the meta positions. It is not clear which position is more favourable. Replacing the six membered ring with a five membered pyrrolidine ring does not increase activity. Only 4 compounds contain the sulphonamide linker, it is therefore difficult to come to a firm conclusion; these preliminary results suggest that replacing the di-keto linker of the oxamide series has little effect on activity.

The affinity of compound KB2_61D for FKBP12 was further investigated by a variety of biochemical and biophysical techniques.

Determining the specificity of KB2_61D for FKBP12

A PPIase assay showed that compound KB2_61D competes with a preferred substrate for the active site of FKBP12; K_d = 2.1 ± 0.8 μM at 4 ^oC. This is in general agreement with the result of the TDF assay. The graph of the background corrected rate of turnover of substrate (0.5 – 1.5 s after addition of substrate) as a function of concentration of inhibitor is shown in Figure 7-20 (A). The experiment was repeated six times.





(D)

Assay	K_d *(μM)*
PPIase **(4 °C)**	2.1 ± 0.8
ITC **(25 °C)**	6.6 ± 0.7
TDF **(55 °C)**	22.2 ± 2.6
TDF extrapolated (ΔH_l) **(25 °C)**	6.22 ± 1.3
ESI-MS (LCQ™DECA) **(50 °C)**	26 ± 13

Figure 7-20 Estimating the K_d for KB2_61D and FKBP12

(A) Inhibition of PPIase activity of FKBP12 by KB2_61D. The initial velocity, V_0 (μM s^{-1}), of FKBP12 catalysed *cis–trans* isomerisation of the tetrapeptide substrate sALPF-pNA is plotted against increasing concentrations of the inhibitor, standard errors for six repeats. The solid line is a best fit to equation 20. (B) ITC thermogram for the interaction between KB2_61D and FKBP12. The experiment was repeated in triplicate (blue, green and black points), error is the standard deviation for 3 separate experiments. Integrated heat data was corrected for heat of ligand dilution and fit to a standard 1:1 binding model, red curves (MicroCal Origin software, V4.0) (C) ESI-MS titration data for KB2_61D and FKBP12. The experiment was repeated in triplicate, error bars are the standard deviation for 3 separate experiments. The solid lines are least squares fit to equation 12. (D) Table of K_d estimated by the different biochemical and biophysical methods.

ITC:FKBP12-KB2_61D complex

The analysis of the ITC data is complicated because the sample of KB2_61D is a diastereotopic mixture. The ITC thermogram for the titration of KB2_61D is shown in Figure 7-20 (B). K_d was estimated to be 6.6 ± 0.7 μM at 25 ^oC making the assumption that all of the ligand in the sample is the active isomer. Modelling studies suggest that not all forms are equally active (section 4.4.4). If the racemic pipecolate reagent consisted of a 1:1 mix of S and R forms and the synthesis did not occur by an steriospecific mechanism the product would be composed of 3 diasteromers: SR (50 %), RR (25 %) and SS (25 %). The activity of each diasteromer is open to conjecture. The modelling study described in 4.4.4 inferred that the activity is likely to decrease in the order: SR > SS >> RR. This would suggest that 50 % of the ligand sample was the most active form with 25 % being a slightly less active. Table 7-2 examines the effect on the ITC model if the compound is considered to be 100 % active and 50 % active. These calculations do not contradict this hypothesis that the sample is less than 100 % active. The activity of the protein sample was determined to be > 95 % by PPIase assay.

Concentration of active KB2_61D	**Stoichiometry N**	**ΔH_{ITC}** kcal·mol^{-1}	**ΔS_{ITC}** cal·K^{-1}·mol^{-1}	**K_d (μM) (25 ^oC)**
500 μM (100 % active)	1.68 ± 0.05	-4.58 ± 0.18	8.0 ± 0.8	7.7 ± 0.8
250 μM (50 % active)	0.89 ± 0.06	-11.66 ± 1.12	-15.4 ± 4.1	6.6 ± 0.7

Table 7-2 ITC for the FKBP12-KB2_61D complex

ITC is sensitive to the concentration of active ligand. Integrated heat data was corrected for heat of ligand dilution (Figure 7-20(B)) and fit to a standard 1:1 binding model (MicroCal Origin software, V4.0).

ESI-MS:FKBP12-KB2_61D complex

The complex of FKBP12 and KB2-KB2_61D had an apparent mass of 13650.4 ± 1.89 Da, this was an addition of 368.7 Da. KB2-KB2_61D has a mass of 368.4 Da, from this one can infer that the complex is a 1:1 ratio of protein to ligand. Figure 7-20 (C) shows the calculated concentration of complex is plotted against the concentration of inhibitor (10 μM FKBP12, 0 – 40 μM KB2_61D). The least squares fit of the data to equation 12 gave a K_d = 26 ± 13 μM, this is in general agreement with the data from ITC and TDF assay. The experiment was performed with a capillary temperature of 50 ^oC. KB2_61D ionised well in electrospray and caused suppression of the protein signal at high ligand concentration. It was not possible to collect data for a concentration of ligand above 40 μM as the protein signal to noise ratio was too low.

Effect of KB2-KB2_61D on the growth of *C.elegans*

Compound KB2_61D (500 μM) was included in the growth media for two strains of *C.elegans*; N2 and TP83. N2 is a wild type worm and TP83 an fkb-3 triple mutant. Fkb-3 is PPIase homologous to mammalian FKBP9 and falls in the secretory pathway class of fkbs (Bell *et al.*, 2006). Treated and untreated worms were compared after 6 days growth, Figure 7-21. Adult worms are typically 1 mm in length. The N2 wild-type worms treated with KB2-KB2_61D showed reduced growth and small vacuoles in the tail region at concentrations above 100 μM. The TP83 worms were more affected; treated worms had vacuoles in the tail and head regions. Organ development was less pronounced in the treated worms. These experiments were carried out by Dr. Victoria Butler (Glasgow University).

18 cyclophilins and 8 FKBPs were identified in the genome of *C. elegans* (C.Elegans sequencing consortium, 1998). There is believed to be a high degree of functional redundancy among PPIases in *C. elegans*. Cyclosporin A and FK506 have been shown to have anti-nematode effects at high concentrations. This is thought to be the result of the compounds inhibiting an endogenous PPIase, rather than through a mechanism involving inhibition of calcineurin. The high concentrations required are thought to be the result of the large number of immunophilin isoforms (Page *et al.*, 1995). The effect of KB2_61D on the phenotype of *C. elegans* was similar to that seen with cyclosporin A or a dimedone derivative (compound 7) (Yang *et al.*, 2007). KB2_61D does not contain a domain that binds to calcineurin; the effect of KB2_61D cannot be a result of calcineurin inhibition and may be due to PPIase inhibition.

Figure 7-21 The effect of KB2_61D on two strains of *C.elegans*

Representative light micrographs of sexually mature worms treated with 100 μM KB2-KB2_61D for 6 days. The worms exhibited growth defects that were consistent with the inhibition of peptidyl-prolyl isomerase activity. The FKB3 triple mutant worms showed more effects than the wild type. The treated worms show poor organ development, reduced size and vacuoles in the tail region (TP83 worms also show vacuoles in the head region). These experiments were carried out by Dr. Victoria Butler (Glasgow University).

7.8.3 **The SL series**

The results of the TDF assay on the SL series are shown in Figure 7-22 Results are grouped by chemical similarity (the clustering method is described in section 7.8.2). The most active compounds were SL6 and SL8; these compounds gave a shift in mid-point melting temperature, ΔT_m > 1.5 ^oC. Table 7-3 shows ΔT_m and K_d estimated at the melting temperature, T_m, (calculated using equation 23, section 6.3.4). K_d were extrapolated to 25 ^oC using 3 putative values of ΔH_L (equation 24).

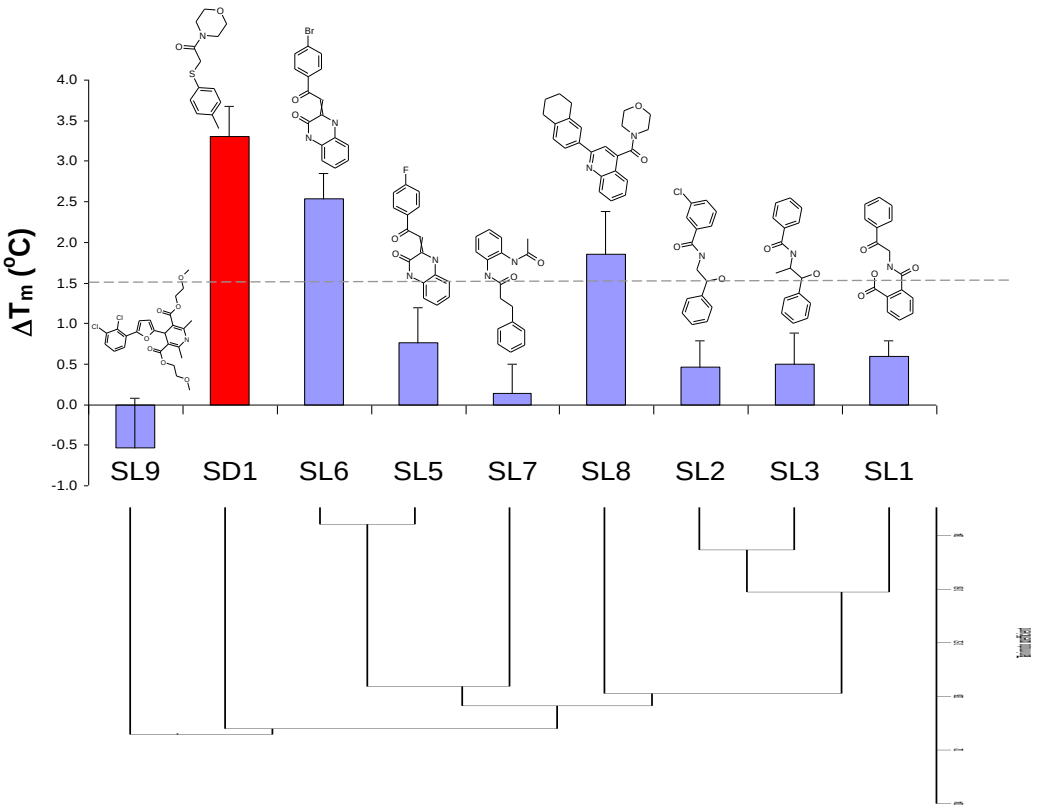


Figure 7-22 TDF assay results for the SL series

TDF assay results for the SL series. 10 μM FKBP12, 250 μM ligand/SD1; 10x SYPRO orange, 1 % methanol, 0.83 % DMSO, 30 mM ammonium acetate; 40 mM Bis-Tris, pH 7. Standard deviation is shown for n = 3 (three separate experiments). The cluster dendrogram is shown to illustrate similar compounds.

Ligand	*ΔT_m* (^oC)	*K_d (μM)* (T_m)	*K_d (μM)* ΔH^a = -2.0 kcal·mol^{-1} (25 °C)	*K_d (μM)* ΔH^b = -5.0 kcal·mol^{-1} (25 °C)	*K_d (μM)* ΔH^c = -7.0 kcal·mol^{-1} (25 °C)
SL8	1.85 ± 0.53	**127**	97	65	49
SL6	2.53 ± 0.32	**98**	74	49	37

Table 7-3 K_d apparent for the two most active members of the SL series

The apparent equilibrium dissociation constant is shown at the melting temperature and at 3 putative values of ΔH_L.

SL1, SL2, SL3 and SL7 were seen to fly in complex with FKBP12 under standard screening conditions on the LCQ^{TM}DECA (ThermoQuest), Figure 7-23. It is hard to discriminate between the ligands from this data or estimate affinity. Only SL1 was seen to fly in complex with FKBP12 on the ZMD (Waters). Ratios of complexed and uncomplexed protein are indicative of ligands in the 100s of μM range. There is no evidence for SL9 binding to FKBP12; this result is consistent with the data from the TDF assay. The TDF was repeated for compounds that dissolved in methanol without the addition of DMSO. This gave a similar trend but larger T_m shifts. This was consistent with DMSO competing for the active site of FKBP12 in the assay where it was included.

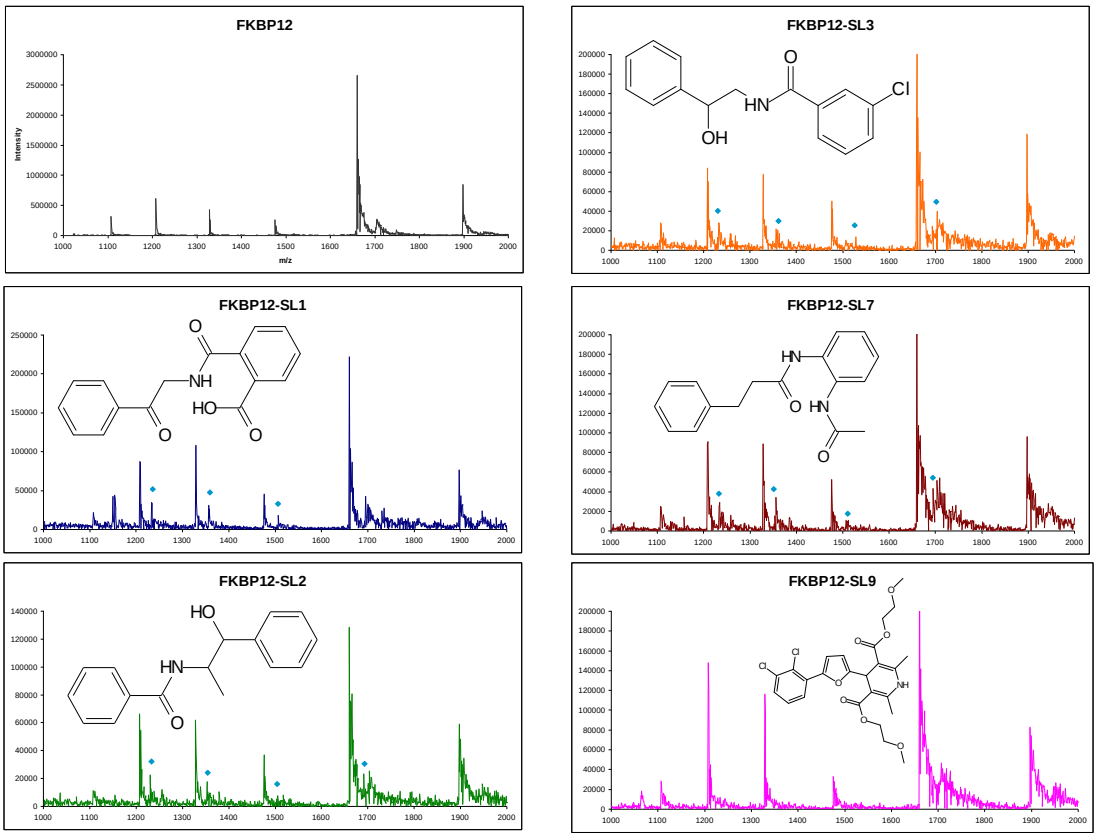


Figure 7-23 ESI-MS spectra for FKBP12 and selected members of the SL series

Representative ESI-MS spectra for FKBP12 and SL1, SL2, SL3, SL7, SL9 (1000 to 2000 m/z). A spectrum of free FKBP12 is shown for comparative purposes. 10 μM FKBP12, 100 μM ligand dissolved in 10 mM ammonium acetate, pH 6.8; 10 % methanol (v/v). Blue diamonds mark peaks corresponding to FKBP12 in complex with ligand.

7.8.4 The Specs library

It has only been possible to test the 162 compounds in the specs set twice. The results presented here are of a preliminary nature as it has not been possible to confirm hits by an alternative assay. Of the 162 compounds 43 (~26 %) compounds show a + ΔT_m > 2 ^oC. This would equate to a K_d < 120 μM (equation 23). If time had allowed the compounds would have been tested a third time and a blank run to test for fluorescence. Melt curves were visually examined to see if there was evidence of a characteristic protein melt in the presence of an inhibitor. Fluorescent compounds typically show high fluorescence at 20 ^oC that often shows thermal quenching as the assay progresses. The positive controls used in the TDF assay gave an extremely consistent ΔT_m; fluorescence compounds or those with solubility problems show much less consistency. A typical T_m for 5 repeats of 250 μM KB2_61D and 15 μM FKBP12 is 55.6±0.3 ^oC. Examining the thermograms would suggest that 28

compounds do not show significant evidence of fluorescence or insolubility and are worth further investigation. Of these 28 compounds, 7 contain a fragment with two rings linked by a sulfonyl group. Compounds containing this pharmacophore were selected from the Organon collection using a BCI fingerprint, the pharmacophore was known to be present in FKBP12 inhibitors reported in the literature. The virtual screening program that showed the highest hit rate was UFSRAT at 26 %. Selecting compounds using a BCI fingerprint (queries were fragments present in a known inhibitor) gave a hit rate of 18 %. The unavoidable presence of 2.5 % DMSO in the TDF screen for this compound collection would have reduced the capacity of the assay to detect low affinity ligands. DMSO binds to the active site of FKBP12 and there would always have been a degree of competition between an inhibitor and DMSO at 2.5 % DMSO.

If time had allowed the next screening step for the hits in the TDF assay would have been purchase the top 30 compounds in solid form and dissolve them in methanol. Methanol is the preferred solvent for ESI-MS and SPR. Compounds with a K_d of less than 30 μM in an ESI-MS or SPR titration would have been tested using ITC as the next step in determining the affinity for FKBP12.

7.9 Discussion and conclusions

The TDF assay has proved a useful assay to test 100s of compounds in a reasonable timescale. It is limited in its ability to discriminate between ligands with a K_d > 100 μM. This is in large part due to the assay estimating K_d at the melting temperature (> 50 ^oC) and the requirement for thermodynamic data to transform the affinity estimation to 25 ^oC (see Figure 6-9). However, lack of sensitivity in the high micromolar affinity range does not distinguish it from most biophysical techniques. To date, only NMR and x-ray crystallography routinely pick up ligands in the high micromolar to millimolar affinity range at moderate screening concentrations. TDF does however have the advantage of being tolerant to low concentration of DMSO. This feature was invaluable for testing the compounds of poor aqueous solubility selected by the docking programs. Low affinity compounds might however be missed due to competition with DMSO.

Virtual Screening Program	*Methodology*	Hit rate (%)[1]
UFSRAT	Ligand based	25
Σ(LIDAEUS score + FlexX score)	Docking	17
Bayesian	Bayesian machine learning	17
LIDEUS score	Docking	7
FlexX score	Docking	7

Table 7-4 Hit rates for different docking programs

[1] A hit was defined as having a predicted $K_d < 120$ μM at T_m in the TDF assay.

Within the time constraints of a PhD it is difficult to acquire enough experimental data to make meaningful comparisons of the enrichment rates for different virtual screening programs. A preliminary analysis would suggest that UFSRAT performed well and that a good score from an arithmetic addition of scores from LIDAEUS and FlexX was a better predictor of activity than a good score in a single program, Table 7-4. In the TDF screen of the Specs green collection 28 compounds had an estimated $K_d < 120$ μM at T_m.

This study suggests that the sensitivity of the TDF assay is a limiting factor in fragment screening. Nonetheless, a fragment with micromolar affinity was discovered with chemical similarity to small molecules known to bind to FKBP12 (F65).

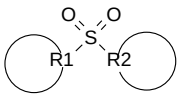


Figure 7-24 Common pharmacophore seen in fragment screening experiments

A common sub-structure in screening hits was a heteroaromatic ring linked by a sulfonyl group to a heterocycle. This is a sub-structure seen in nanomolar affinity ligand in the literature and occupies the most buried region of the active site (section 1.11.6 and Appendix 3).

The 3 fragments contained a substructure consisting of a heteroaromatic ring linked by a sulfonyl group to a heterocycle, Figure 7-24. This sub-structure was also seen to have a small $+\Delta T_m$ in KB series (3 compounds in the bis-sulphonamide subset; Figure 7-11). In the Specs green collection 7 of the 28 compounds that gave a $+\Delta T_m > 2^oC$ contained this sub-structure (best hit K_d ~5 μM). Although it is worth noting that these particular Specs or KB series compounds were synthesised or selected from a compound collection using knowledge derived from the literature search for this project. They can perhaps be considered confirmation of the similarity principle rather than novel inhibitors.

The KB series of compounds confirmed the observation from the literature that a substitution of –CO-R in the ortho position on a heterocyclic ring (that takes the place of proline of the natural substrate) is important to affinity. This group is present in the majority of nanomolar inhibitors of FKBP12 (Figure 1-10). The active site of FKBP12 is a broad, shallow pocket and a hydrogen bond between a carbonyl group and the main chain N of Ile56 is a common feature of inhibitors. The best inhibitor in the KB series, KB2_61D, had an affinity of 6.6 μM (25 ^{o}C by ITC); it inhibited the turnover of the preferred substrate for FKBP12 in the PPIase assay. This compound was a diastereotopic mixture, the modelling data suggests that one of the 3 diastereomers is more active and that it is probable that the single (S),(R) form may have an affinity for FKBP12 in the high nanomolar range.

8 Chapter 8 Human Cyclophilin A – Testing the KB Series

8.1 Introduction

The major part of this study investigated protein-ligand interactions for FKBP12. The KB series were designed to target two immunophilins; FKBP12 and hcypA (section 7.7.2 describes the design of the series). Experimental studies on hcypA were only carried out for the KB series. The full KB series was screened against hcypA using the TDF assay. Selected compounds were examined by x-ray crystallography (co-crystallisation and crystal soaking). The crystallisation conditions for hcypA are well established. The aim of the crystallisation experiments was to detect binding for low affinity ligands (K_d in the mM range).

8.2 Materials and Methods

8.2.1 Protein Production

Recombinant untagged human cyclophilin A (hcypa) was expressed in *E.coli* (BL21 Star (DE3), Figure 8-1) and purified under native conditions following the protocol described by Wear (Wear *et al.*, 2005). Pure protein yields were in the order of 5 $mg \cdot L^{-1}$ culture. hcypA was judged to be greater than 95 % purity by SDS-PAGE, Figure 8-2.

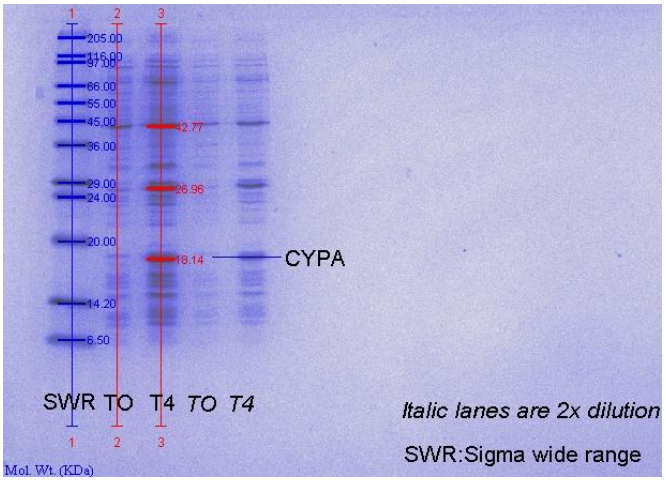


Figure 8-1 Expression of hcypA

15 % acrylamide SDS-PAGE gel under reducing conditions showing overexpression of hcypA in *E.coli* (recombinant strain BL21star™) grown in 2xTY media at 37 °C with 260 rpm shaking. Cells were induced after reaching an optical density of A_{600nm} ~ 0.6 with 1 mM IPTG. Lanes TO and T4 show the soluble fraction from before and after 4 hours induction with IPTG. Italic lanes are a 2x dilution. Molecular weight marker (MW) Sigma wide range. Molecular weight markers are annotated with MW (kDa) in blue, whole cell extract in red. hcypA ran at 18.1 kDa in SDS-PAGE (15 % acrylamide).

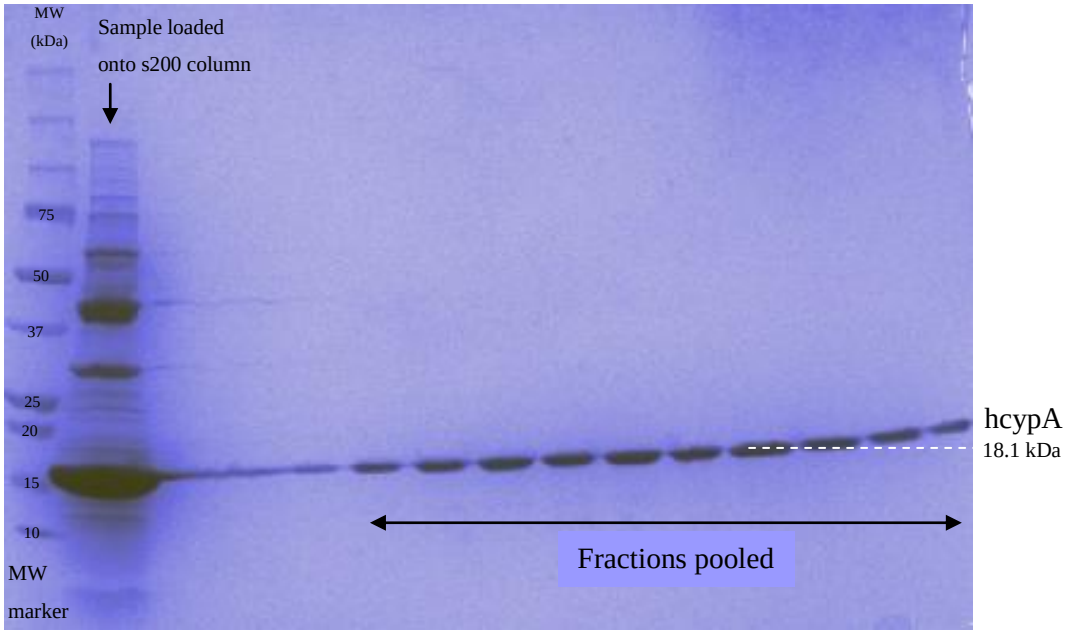


Figure 8-2 hcypA after purification by gel filtration

15 % acrylamide SDS-PAGE gel under reducing conditions showing hcypA after the final step of purification. The cell extract (pH adjusted to 6.8 in the buffer used for the cation exchange column) was applied to an HiPrep 16/10 SP FF (GE Healthcare) cation exchange column pre-equilibrated with 50 mM HEPES, pH 6.8; 1mM DTT, 2.5 mM EDTA, 1 mM sodium azide. The column was eluted with a gradient of 0 to 400 mM sodium chloride. The pooled fractions were concentrated to 1.5 ml and passed over a HiPrep 16/60 s200 gel filtration column (GE Healthcare) pre-equilibrated with 10 mM sodium phosphate (pH 7), 100mM sodium chloride, 0.5 mM DTT, 0.5mM EDTA and 1 mM sodium azide. MW marker: Precision Plus Protein Standard (Bio-Rad).

8.2.2 **Protein characterisation (hcypA)**

The concentration of hcypA was measured by absorption at 280 nm using an extinction coefficient of 8490 $M^{-1} \cdot cm^{-1}$ on a nanovue instrument (GE Healthcare). The molecular weight of hcypA is 18012 Da. This was confirmed by ESI-MS performed under denaturing conditions following the protocol described in section 3.5.1, Figure 8-3. The protein flew with an apparent mass of 18147 Da and was the dominant species in the sample; this is consistent with other studies.

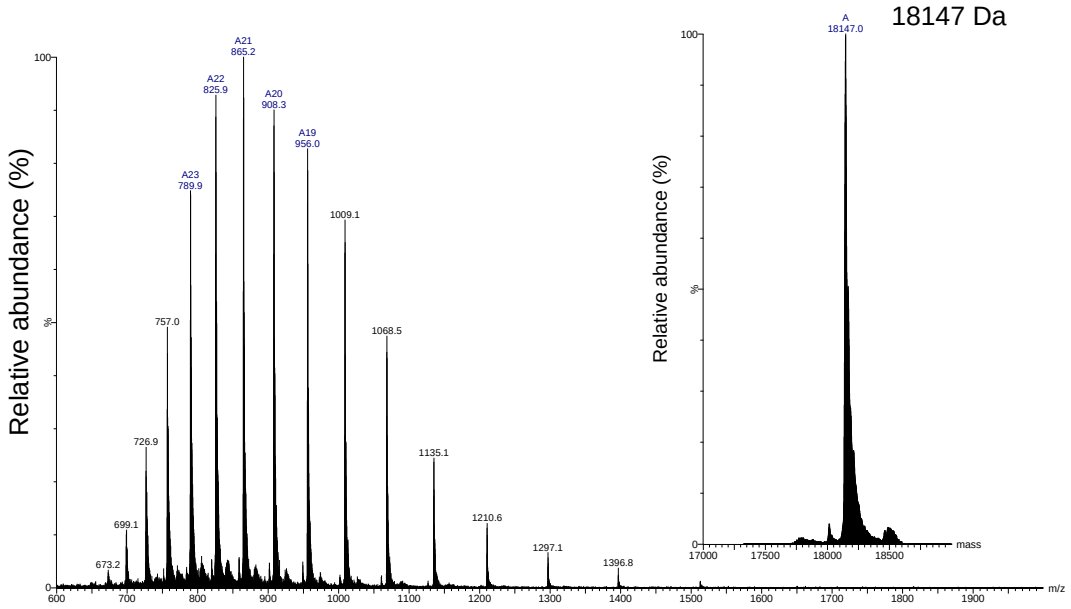


Figure 8-3 ESI-MS spectrum of hcypA

20 μM hcypa was flown in 20 mM ammonium acetate, pH 2.5; 2 % Formic acid; 10 % methanol (v/v) flown under default conditions. The inset diagram shows the transformed spectrum labelled with the mass of hcypA, the dominant species. The spectra were background corrected smoothed and transformed using MassLynx V4.1 (Waters).

8.2.3 **Screening the KB series by TDF assay**

The TDF assay was performed as described in sections 6.3.4 and 7.7.2. The only significant differences to the method were the concentration of hcypA and the positive control. hcypA (18 kDa) is a larger protein than FKBP12 (13 kDa). Scouting experiments (described for FKBP12 in section 6.4.4) showed that 5 μM hcypA was sufficient to give a large enough fluorescent change on thermal denaturation of the protein. Cyclosporin A (CsA) was used as a positive control at a concentration of 8 μM. CsA was not soluble beyond 10 μM in aqueous buffer.

The 20 compounds in the KB series were tested in parallel on a single plate. 3 positive controls (CsA) and 3 negative controls were included on each plate. The final solution contained 5 μM hcypA, 250 μM Ligand/8 μM cyclosporin (positive control); 30 mM bis-tris, pH 7; and 40 mM ammonium acetate; 0.25 % methanol and 10x SYPRO orange. The experiment was repeated in triplicate in separate experiments. A fluorescence control was also included for each compound on the plate. Plates were incubated at room temperature for 15 minutes prior to the start of the assay. Data was processed using the methods described in section 6.3.4.

8.2.4 **Crystallisation of hcypA**

The crystallisation and crystal soaking protocol was modified from the method described by (Yang *et al.*, 2007). hcypA was exchanged into 20 mM HEPES, pH 8; 100 mM sodium chloride; by PD10 rapid desalt column (GE Healthcare) and concentrated to ~ 26 $mg \cdot ml^{-1}$ by spin concentration (Vivaspin 500, MW cut-off 5 kDa). Crystals were grown by vapour diffusion in 24 well Linbro plates (Molecular dimensions) by the hanging drop method at 277 K. The initial 5 μl drop consisted of 13 $mg \cdot ml^{-1}$ hcypA, 10 mM HEPES, 50 mM Tris, pH 8; 100 mM sodium chloride, 14 to 16 % PEG8000 (v/v). The well solution consisted of 100 mM Tris, pH 8, 100 mM sodium chloride, 28 to 32 % PEG8000 (v/v). Crystals grew within a few hours and reached a maximum size after ~3 days. Crystals formed with and without the presence of 5 % DMSO (Kontopidis *et al.* (2004) added 5 % DMSO to aid crystallisation which they then soaked out).

The crystals were soaked by transferring them into pre-equilibrated drops (14 hours) containing increasing concentrations of ligand dissolved in the well solution (25 %, 50 %, 75 %, 100 % (v/v) saturated solution of ligand; otherwise similar to the well solution). The crystal was soaked in the final solution for 3 hours. The success rate for the soaking

experiments was variable; some ligands caused the crystal to crack at a low concentration. DMSO was not included after it was found to be unnecessary for crystallisation. Co-crystallisation was tried for ligand KB2_61D by adding 10 mM ligand to the drop ([ligand] > 10x [protein]). Crystals formed in drops containing 16 to 18 % PEG8000 in this series of experiments.

The crystals were flash frozen in liquid nitrogen after soaking in a cryoprotectant solution similar to the final soaking/co-crystallisation solution with the addition of 22 % glycerol (step-wise soaking in increasing concentrations of glycerol).

Data was collected at SRS, Daresbury (λ = 0.95 Å) or SRS, Diamond (λ = 0.97 Å) and processed with the CCP4 suite of programs V6.0.2 (CCP4, 1994). Frames were read and integrated with MOSFLM (Powell, 1999) and scaled using SCALA (CCP4, 1994). Molecular replacement was performed using MOLREP (Lebedev *et al.*, 2008). An x-ray structure of hcypA from a complex containing the di-peptide ligand Ala-Pro (2CYH.pdb, (Zhao and Ke, 1996)) was used as the model for molecular replacement. This structure has 100 % sequence identity with the recombinant hcypA used in this study. Structural refinement was carried out using Refmac5 (Murshudov *et al.*, 1997). Output from the programs were viewed and structural models altered for rounds of structural refinement using the model building program COOT (Emsley and Cowtan, 2004).

8.3 Results

8.3.1 Affinity of hcypA for cyclosporin A

The immunosuppressant CsA, was used as the positive control for binding studies with hcypA. CsA is an undecapeptide that binds to hcypA with 1:1 stoiciometry with low nanomolar affinity. hcypA exhibited an affinity for CsA in general agreement with the literature. The K_d of CsA for hcypA has been reported from 1.6 to 46 nM by a variety of different assays at a range of temperatures (Wear *et al.*, 2005; Wear and Walkinshaw, 2006). Wear et al. (2005) compare K_d from the literature.

Equation 22 was fitted to the raw fluorescence data using the method described in section 6.3.4 (inset diagram Figure 8-2 (A)). ΔH_U= 12.72 $kcal^{-1}\cdot mol^{-1}$, ΔC_{pU} = 6180 $cal^{-1}\cdot K^{-1}\cdot mol^{-1}$. Figure 8-4 (B) shows the change in mid-point melting temperature on addition of 10 μM CsA to 5 μM hcypA ($+\Delta T_m$ = 5.7 ± 0.3 ^oC). It was interesting to note that the background fluorescence for native hcypA was much lower than that of FKBP12. This could be rationalised by hcypA having fewer hydrophobic surface residues than FKBP12.

The affinity of CsA was calculated at the mid-point melting temperature (T_m) using equation 23. K_d was extrapolated to 25 ^oC using thermodynamic parameters calculated by Wear and Walkinshaw (2006) and equation 25. K_d for the hcypA-CsA complex are summarised in Table 8-1.

Ligand	***$K_{d(TDF)}$ (nM) T_m***	***$K_{d(TDF)}$ (nM) (Extrapolated to 25 ^oC)***	***Literature K_d (nM)***
CsA	240 ± 50	7.9 ± 2.0	1.6 – 46 (Wear *et al.*, 2005)

Table 8-1 Affinity of CsA for hcypA measured in the TDF assay

Affinity measured by TDF was transformed to 25 ^oC using equation 25; ΔH_L= -14.600 $kcal^{-1}\cdot mol^{-1}$, ΔC_p = -870 $cal^{-1}\cdot K^{-1}\cdot mol^{-1}$ were taken from the literature (Wear and Walkinshaw, 2006). Standard deviations are shown for n = 3.

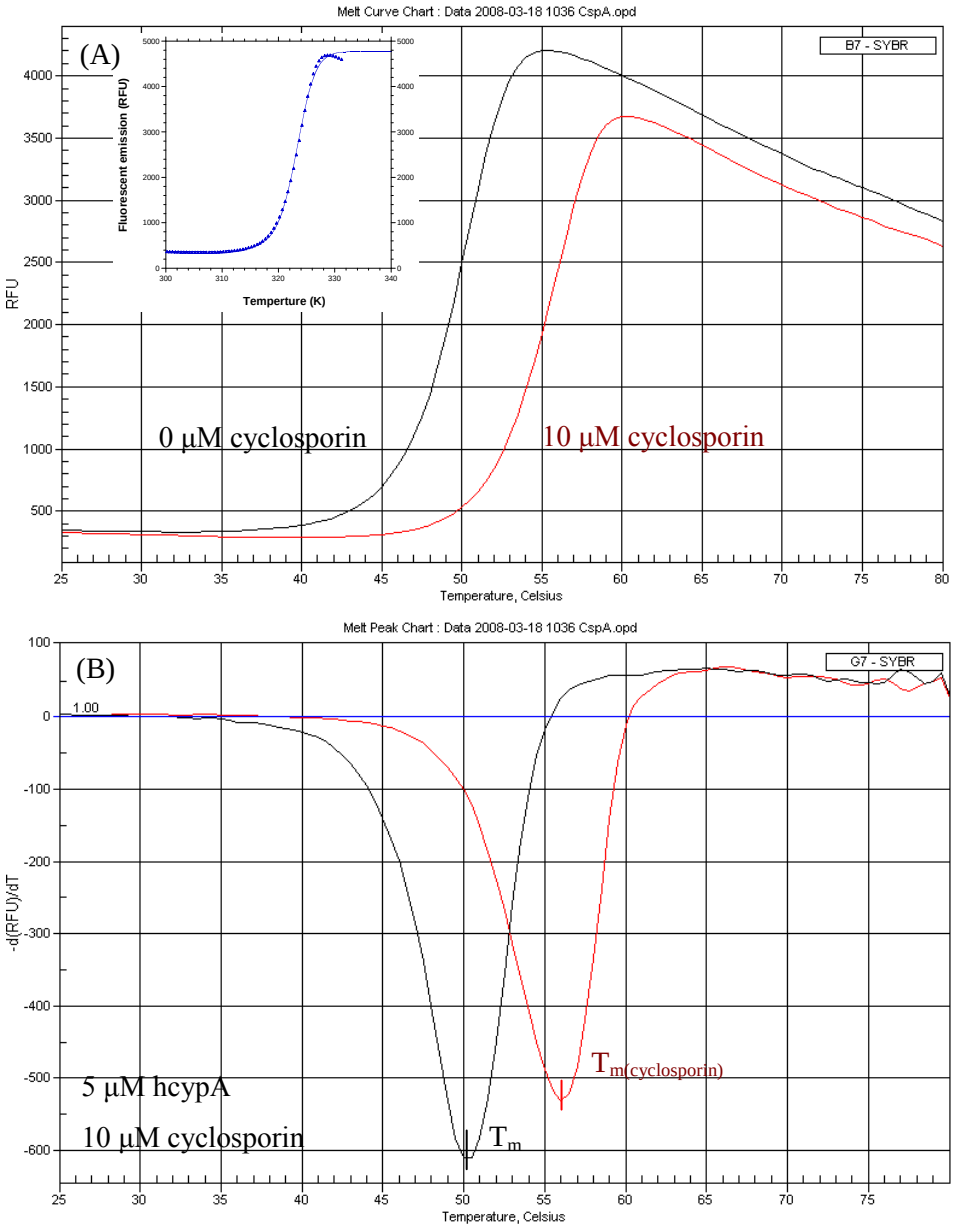


Figure 8-4 TDF thermograms for hcypA and cyclosporin

(A) Typical raw TDF data. Melting curves are shown for 5 µM hcypA; 10 µM (red) cyclosporin; 10x SYPRO; 1% methanol; 50 mM ammonium acetate, pH 7.0. Thermal denaturation occurs between 30 and 60 ^{o}C. Inset curve in blue is the best fit of data to equation 22 using the non-linear regression program Kaleidagraph (V4.03, Synergy software) (B) –dF/dT for graph (A). T_m is labelled for the protein and after addition of 10 µM cyclosporin.

8.3.2 **KB series**

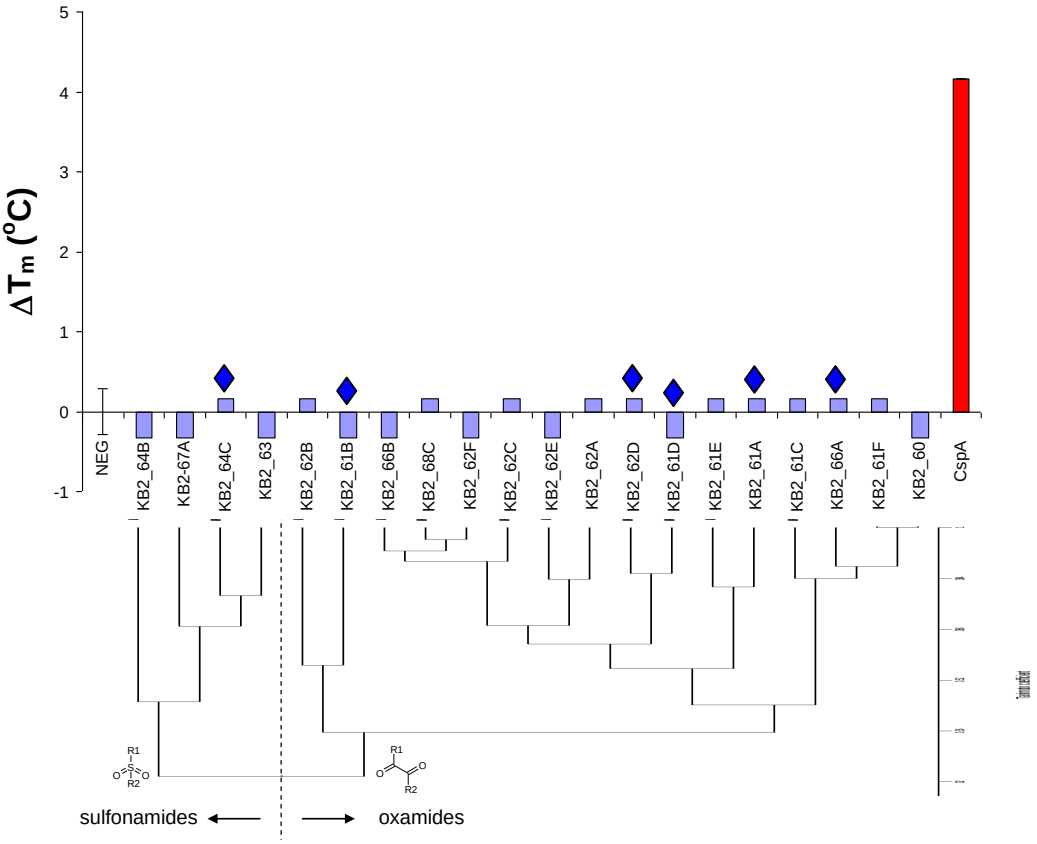


Figure 8-5 Screening the KB series by TDF assay (hcypA)

5 µM FKBP12; 8 µM (red) cyclosporin/500 µM test compound, 10x SYPRO; 0.25 % methanol, 30mM ammonium acetate; 40 mM Bis-Tris, pH 7. The red bar represents the positive control CsA. Blue diamonds indicate that the ligands were tested in the crystallographic study.

None of the KB series was found to bind to hcypA. Results of the TDF assay are shown in Figure 8-5. Results for the positive control CsA were in agreement with the literature (Table 8-1).

Crystallisation

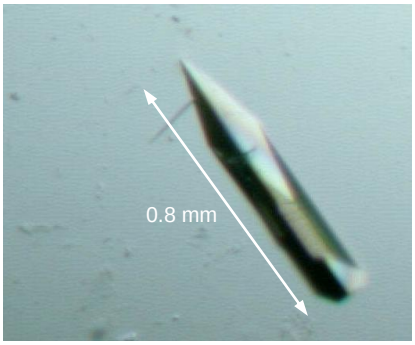


Figure 8-6 hcypA crystal grown in 16 % PEG8000, pH 8
The maximum dimension of this crystal was 0.8 mm.

X-ray crystallisation has the potential to detect the binding of ligands in the mM range if they are sufficiently soluble under crystallisation conditions (Hartshorn *et al.*, 2005). The crystallisation of hcypA proved to be relatively straightforward. Data was collected for hcypA in the presence of saturating concentrations of 5 of the KB series. These were solved to a resolution between 1.8 and 2.7 Å in space group $P2_12_12_1$. None of the ligands tested in the soaking and cocrystallisation experiments were found in the active site of hcypA. Results are summarised in Table 8-2. A typical hcypA crystal is shown in Figure 8-6.

Summary of soaking/co-crystallisation experiments – hcypA and the KB series

Compound	*Method*	*Diffraction*	*Results*
KB2_61D	Co-crystallisation 1:12 protein:ligand ratio	YES	no ligand in active site
KB2_62D	3 hour soak in sat sol^n	YES	no ligand in active site
KB2_61A	3 hour soak in sat sol^n	YES	no ligand in active site
KB2_61B	3 hour soak in sat sol^n	YES	no ligand in active site
KB2_66A	15 minute soak in sat sol^n	YES	no ligand in active site
KB2_64C	15 minute soak in sat sol^n	YES	Crystal diffracted, damage to crystal on soaking, no data set taken

Table 8-2 Summary of crystallisation experiments

8.4 Discussion and conclusions

In the study of Kontipidis *et al*. KB2_68C was crystallised in complex with hcypA (Figure 7-8, (Kontopidis *et al.*, 2004)). There was no evidence that this bound in the TDF assay performed in this study. This probably not a surprising result as the affinity of this compound for hcypA was in the order of 25 mM and this is beyond the sensitivity of the TDF assay. Unfortunately the soaking experiments were not successful for this ligand and results were not reproduced. No other members of the KB series were found to bind to hcypA with TDF or x-ray studies. The region of the active site that binds KB2_68C is a small pocket adjacent to Phe113; these results would suggest that substitutions to the piperidine ring ortho to the di-keto linker make the ligands tested too bulky to occupy this site. A substitution of –CO-R in the ortho position on a heterocyclic ring is present in all most inhibitors of FKBP12 (Figure 1-10). The active site of FKBP12 is a much broader pocket and a hydrogen bond between a carbonyl group and the main chain N of Ile56 is a common feature of inhibitors.

9 Chapter 9 Summary and Future Work

The study explored a series of biophysical assays to determine the best screening methodology to test hits from high-throughput virtual screening programs. The aim was to inhibit the human immunophilin FKBP12, a target for the treatment of neurodegenerative diseases. The conclusions from this study can be divided into four main areas.

1. **Studying protein-ligand interactions by ESI-MS.**
   A detailed evaluation was carried out to test the suitability of ESI-MS for estimating ligand affinity in FKBP12-ligand complexes. There was good correlation between the rank order of affinity in ESI-MS and in solution phase experiments. Collision induced dissociation experiments, with protein-ligand complexes, showed that there was a correlation between the cone voltage that gave 50 % dissociation of the complex (VC50) and the enthalpic contribution to the Gibbs free energy of binding in three different protein systems (Chapters 3 and 4).

2. **Extending the sensitivity of ESI-MS – The charge-state ghost**
   A novel method was investigated that extended the sensitivity of ESI-MS to detect very weak inhibitors of FKBP12 by examining changes to the charge-state distribution of protein molecular ions (Chapter 5).

3. **A comparison of different biophysical methods to characterise protein-ligand binding affinity**
   A test set of inhibitors of FKBP12; with diverse structure and a wide range of affinity was complied to assess the sensitivity and throughput of a series of established biophysical assays. These assays were compared to results from the ESI-MS; there was good agreement between the different biophysical techniques (Chapter 6).

4. **Screening for FKBP12 inhibitors**
   Hits from virtual screening, a fragment library and compounds designed to target multiple immunophilins were tested in the laboratory using biophysical methods Over 300 compounds were screened in the laboratory and 34 inhibitors identified in preliminary screens. (Chapter 7 and 8).

9.1 Studying protein-ligand interactions by ESI-MS

9.1.1 Aims

The first aim of this section of the study was to fly the six-histidine construct of FKBP12 in ESI-MS. The second aim was to estimate the equilibrium dissociation constant for a complex from a comparison of signal intensity between free protein (protein ion) and protein in complex with a ligand (complex ion).

9.1.2 Major results and conclusions

The six-histidine construct of FKBP12 flew well on all instruments tested; the protein exhibited a narrow charge-state distribution dominated by the 8+ charge-state. Between 5 and 10 μM protein in 10 mM ammonium acetate, pH 6.8; 10 % methanol (v/v) gave a reproducible spectrum and good signal to noise ratio.

Single point screening experiments in ESI-MS were information rich. Experiments carried out at a single ligand concentration gave data on the mass of a ligand, its relative binding affinity/stoichiometry and also screened for protein stability. Ligand or pH conditions that induced protein unfolding in solution gave a much broader charge-state distribution than correctly folded protein.

A method was developed for calculating K_d from an ESI-MS titration experiment. These experiments gave an estimate of K_d close to the value measured in solution for ligands with low nanomolar affinity for FKBP12. The lowest affinity ligand in the test set seen to fly in complex with FKBP12 had a solution K_d of 157 μM. K_d were consistently over estimated by the titration method. This was attributed to collision induced dissociation of the complex. Nonetheless, rank order of affinity was similar to that measured in solution. Setting cone voltage in ESI-MS was always a compromise between sensitivity and good signal to noise ratio. Cone voltage dissociation experiments established 15V to be the minimum cone voltage required to screen a series of compounds on the ZQ instrument using standards screening concentrations.

Cone voltage dissociation profiles were used to quantify the relative fragility of a particular complex in electrospray by calculating the cone voltage required to give 50 % dissociation of the complex. This value was defined as the VC50. Studies have suggested there might be a correlation between VC50 and affinity in solution (Daniel *et al.*, 2002). In this study VC50 was a poor indicator of affinity in solution. Comparison of VC50 and ITC data showed that

there was good correlation between VC50 and the enthalpic contribution to the free energy of binding (ΔH_{ITC}). There was also correlation with the number of hydrogen bonds between protein and ligand predicted by x-ray crystallography or modelling studies.

The lack of correlation between VC50 and affinity in solution for FKBP12 has consequences for the use of ESI-MS as a screening technique. The results of this study suggest that complexes with a high component of a favourable free energy change on binding due to the hydrophobic effect are likely to be fragile in electrospray. Some complexes of low affinity ligands might appear as false negatives in a screen. However, this phenomenon could be used to advantage during lead optimisation. Drugs with nanomolar or better affinity for a target generally have favourable contributions from both enthalpic and entropic terms (Carbonell and Freire, 2005). Medicinal chemists often aim to improve the affinity of a drug for its target by the addition of non-polar groups during lead optimisation. The hydrophobic effect leads to an entropic gain on formation of the complex. However, drugs unduly dependent on the hydrophobic effect often have poor solubility. The optimisation of a drug by improving the number of electrostatic interactions between drug and protein is challenging. The aim is to improve affinity by including functionality that increases the number of electrostatic interactions. Polar groups consisting of hydrogen bond donors or acceptors can be added to the compound to make additional electrostatic interactions. Strong electrostatic interactions are dependent on the geometry of residues in the active site and can aid specificity for the target. It is self-evident that a small addition of $-\Delta H$ has a significant influence on the affinity. For example reducing ΔH by around 1.5 $kcal \cdot mol^{-1}$ improves affinity ~10 fold at $25^{o}C$ due to the logarithmic nature of the relationship described by equation 2. ESI-MS might provide a fast route to structure-activity relationships for a similar chemical series. Small changes in structure can often lead to large changes in affinity and it is useful to know what role electrostatic interactions might play (Eckert and Bajorath, 2007). A thermodynamic profile can inform the selection of lead compounds from primary hits.

As a generalisation, biological ligands show a greater dependence on a favourable enthalpic term than most inhibitory drugs (Li *et al.*, 2008). The binding of sugars and DNA to a protein often have an unfavourable entropic term that is counterbalanced by the ligand making numerous enthalpically favourable hydrogen bonds to the protein. Carbohydrate ligands, because of their modular nature, allow for great stereo chemical diversity and hence selectivity. Enzymatically catalysed synthetic routes *in vivo* are evolutionally optimised to deliver the preferred stereoisomer and conformation from very simple building blocks.

Stereo specific synthetic routes in the laboratory can be difficult and time consuming, as it is costly to perform separation. Macrocyclic drugs, such as FK506 and rapamycin, derived from natural products (often fungal metabolites) could be said to form a class of compounds with characteristics of both biological ligands (sugars, DNA and peptides) and small molecule drugs. It has been suggested that macrocycles offer a good compromise between a conformationally pre-organised structure and flexibility. Key fuctional groups are correctly arranged to hit "hot" residues some distance apart on the protein surface with just enough ring flexibility to allow binding without a large entropic loss (Driggers *et al.*, 2008).

The correlation between VC50 and ΔH_{ITC} is interesting as ΔH_{ITC} is not equivalent to the interaction energy between ligand and protein in the gas-phase. The magnitude of ΔH_{ITC} also reflects the enthalpy associated with desolvating the ligand from the aqueous media. Bonds are broken with water when the hydrogen bonding network of the bulk solvent is altered as the ligand is removed from the solvent. The magnitude of the enthalpy of desolvation is ligand dependent. However, the "solvent past" of the complex can't be overlooked. After a macromolecular complex is desolvated in ESI-MS there is no equilibrium between the complex and the binding partners. There are likely to be extremely few binding events compared to the number of ligands dissociating from a complex. The equilibrium in solution is important. The on-rate of the ligand is influenced by the presence of water and hydrophobic effects. After desolvation in ESI-MS, the off-rate is likely to be largely dependent on the enthalpic contribution to binding from electrostatic interactions between protein and ligand.

9.1.3 **Future work**

It would be interesting to extend this study to a larger collection of ligands. In the final series of screening experiments described in Chapter 7 over 28 new ligands appeared as hits in preliminary screens. It would be very interesting to include these in an ESI-MS study. K_d estimated by ESI-MS and VC50 could be compared with ITC and x-ray crystallographic data. Carbonic anhydrase might be an interesting protein to extend this study as many small soluble inhibitors are available over a range of K_d (Krishnamurthy *et al.*, 2008). Carbonic anhydrase has been widely used as a model protein for biophysical studies.

9.2 Extending the sensitivity of ESI-MS – the charge-state ghost

9.2.1 Aim

This part of the study stemmed from the serendipitous observation that charge-state shifts accompanied dissociation of a protein-ligand complex.

9.2.2 Major results and conclusions

Mass spectrometry is regularly used in drug discovery to screen for non-covalent interactions between proteins and ligands. An interaction is defined by the presence of a peak in the electrospray spectrum indicative of protein in complex with a ligand (the complex ion). However, for weakly bound complexes (with K_d in the micromole range) this complex ion is not always observed because of collision induced dissociation of the protein-ligand complex. This technique (of monitoring the complex ion) is limited to relatively tightly bound complexes; as a result promising ligands can be undetectable or incorrectly ranked in a screen. In this study, the results described in Chapter 4 suggested that compounds with a significant entropic component to a favourable free energy change on ligand binding to be particularly fragile in electrospray.

It was that found charge-state shift to be a significantly more sensitive screening technique than the current methodology of identifying protein-ligand complex ions in the electrospray spectrum. Charge-state shifts were observed for three protein-ligand systems on dissociation of a non-covalently bound complex.

9.2.3 The charge-state method for the identification of low affinity interactions

The charge-state method may have particular application when screening small fragment-like molecules which have affinity in the micromolar range but become potential lead compounds after fragment linking or chemical modification. The identification of low affinity ligands by biophysical methods is currently challenging.

To measure a weak interaction the ratio of ligand to protein must be selected to give a sufficient physical response to measure in the assay, this is often limited by ligand solubility if a high concentration of protein is required. ITC has the potential to detect weak interactions but is a protein hungry technique that requires the ligand to be soluble to

relatively high concentrations (mM). It is not routinely used for screening experiments. SPR has been successfully used for small molecule screening for some protein systems but it dependent on the immobilisation chemistry retaining high levels of protein activity (Neumann *et al.*, 2007). It also involves considerable capital outlay for instruments and high running costs for consumables. Immobilising ligands onto beads by means of a chemical linker that allows the ligand to behave in a similar manner to a free ligand in solution facilitates the desirable protein ligand ratios. Combining this with very sensitive fluorescence based methods enables the technique to detect a very wide affinity range. (Hintersteiner and Auer, 2008). However, this method does require a high level of robotic assistance with its associated costs. NMR has been successfully used to detect weak interactions at attainable protein to ligand screening ratios as demonstrated by the work of Stebbins *et al*. It is an attractive alternative methodology albeit with a high capital outlay for instrumentation.

In a university setting ESI-MS uses relatively simple and inexpensive equipment that is often already in place. The technique is not sensitive to coloured or fluorescent compounds that can be a problem in some systems where detection is based on optical absorbance or fluorescence. On a more industrial scale the technology for the automation sample delivery is available (Nanomate^{TM}, Advion Biosciences, Inc.).

9.2.4 **Future work**

It is important to test the charge-state shift hypothesis with a larger test set of diverse ligands. The ligand identified in the screening experiments described in Chapter 7 would provide such a set of compounds for FKBP12. Carbonic anhydrase would also be a useful model system to investigate. Only after a large compound set and a diverse range of targets have been investigated can the rates of false positives and false negatives be compared to other methodologies.

Proof of principle in a nanospray instrument is a vital next step. This would reduce reagent consumption and enable the use of automated sample delivery systems that would be essential for high screening rates.

9.3 A comparison of biophysical methods to characterise protein-ligand complexes

9.3.1 Aims

It was important to compare results from estimation of ligand affinity by ESI-MS with experimental approaches that rely on different physical phenomenon. Compounds from the ligand test set were used to compare different biochemical and biophysical methods. The aim was to choose the best assay for preliminary screening experiments at a single ligand concentration to provide the most active subset for re-testing.

9.3.2 Major results and conclusions

A comparison was made of the advantages of each screening method in terms of sensitivity, richness of thermodynamic data, throughput, reagent consumption and cost. Each method delivered an apparent K_d that must be considered with the inherent assumptions made in the model and the possibility of false positives. Only ITC provides a direct measurement of the enthalpy of a binding interaction. However, it was clear that ITC could not provide all the information required within a short timescale or with the reagents and funding available. It was necessary to design a strategy consisting of a preliminary screen at a single ligand concentration to cut down the number of compounds to be tested by more rigorous methods. A TDF assay was selected to perform the initial screen. The top 20 % were then retested by ESI-MS. This step was dependent on the ligand being soluble in methanol; DMSO was incompatible with this assay. The top hits were tested by ITC.

For ligand with affinity less than 10 μM there was remarkable consistency between different assays (Table 6-6 and Figure 7-20). Perhaps contrary to expectations, estimating K_d by ESI-MS using the established procedure of measuring the relative intensities of the protein and complex ions performed at least as well as other biophysical techniques. The charge-state ghost method described in Chapter 5 proved to be the most sensitive technique at the screening concentrations used in this study. However, this technique is most suited to an initial screen rather than the detailed characterisation of binding.

The ligand test set proved a useful set of compounds to test the different assays. SD1 proved to a particularly useful positive control, as it was more than 100 x more soluble than the nanomolar affinity ligands FK506 and rapamycin. It was also closer in affinity to ligands that might be expected to be discovered in screening (K_d = 6.8 ± 1.2 μM). The enthalpy of

binding of SD1 was measured by ITC (ΔH = -6.1± 0.6 kcal·mol^{-1}). This is a more useful generic ΔH for a small molecule inhibitor of FKBP12 than the potent immunosuppressants FK506 or rapamycin. This makes it particularly useful as a positive control in the TDF assay. The ligand KB2_61D, discovered during the course of these experiments, proved to be a good positive control in the final screening experiments.

9.4 Screening for FKBP12 inhibitors

9.4.1 Aims

The aim was to discover novel non-immunosuppressant ligands for FKBP12. The primary aim was to test hits from virtual screening programs. Hits were tested from LIDAEUS, FlexX, UFSRAT and a Bayesian machine learning method. A fragment library was also tested and a series of compounds designed to target both FKBP12 and hcypA (KB series).

9.4.2 Major results and conclusions

The TDF assay has proved very successful in estimating affinity for low micromolar ligands of FKBP12 and a nanomolar inhibitor of hcypA, the immunosuppressant CsA. It has been a useful assay to test 100s of compounds in a few days. However, it is limited in its ability to discriminate between ligands with a K_d > 100 μM. This is largely due to the need to transform data from 55 ^{o}C to 25 ^{o}C using thermodynamic information (see Figure 6-9). At the present time it is challenging to detect ligands with an affinity > 150 μM by any of the methods used in this thesis. This is a limiting factor in fragment screening. Nonetheless, a fragment with micromolar affinity was discovered with chemical similarity to small molecules known to bind to FKBP12 (F65).

A sub-structure was seen in the Specs green collection that is commonly seen in ligands of nanomolar affinity reported in the literature (present in 7 compounds). This consists of a heteroaromatic ring linked by a sulfonyl group to a heterocycle. This moiety occupies the most buried region of the active site in known inhibitors with crystallographic data (section 1.11.6 and Appendix 3 for references).The testing of the Specs green collection is still a work in progress. Of the 162 compounds tested 28 showed some evidence of activity in the preliminary TDF assay.

The KB series of compounds confirmed the observation from the literature that a substitution of –CO-R in the ortho position on a heterocyclic ring, taking the place of proline of the natural substrate, is important to affinity. This group is present in the majority of nanomolar inhibitors of FKBP12 (Figure 1-10). The active site of FKBP12 is a broad, shallow pocket and a hydrogen bond between a carbonyl group and the main chain N of Ile56 is a common feature of inhibitors. The most active compound in the KB series against was compound KB2_61D with a K_d of 6.6 μM for FKBP12. This compound is in a diastereotopic mixture and a single (S),(R) diastereomer is likely to have affinity in the high nanomolar range.

9.4.3 **Future Work**

It would be extremely interesting to test the 28 hits from the Specs green collection by ESI-MS, ITC and SPR. If hits were confirmed by these alternative methods; the similarity search function of EDULISS 2 (UFSRAT and the descriptor and graph-based methods of Hsin) would make the selection of similar compounds straightforward (Hsin, 2009). At this stage it would be possible to make firmer comment on the relative enrichment rates of the different virtual screening programs. However to date the evidence suggests that all methods show some enrichment, particularly UFSRAT.

10 Appendix 1 – Exploring competition for the active site

Solving the law of mass action to calculate the proportion of a ligand in complex with a single-site binding model and one ligand in solution is mathematically straightforward. Equation 1 describes the law of mass action and equation 12 the solution to the quadratic equation. The situation becomes more complex when two ligands, A and B, are in solution are competing for the same site on the protein (P). The ligand B will displace A from the active site of the protein to a degree that is dependent on the affinity and concentration of A and B. The equilibria between the different species can be described by equations 31 and 32.

$$P + A \underset{K_2}{\overset{K_1}{\rightleftharpoons}} PA \qquad \textbf{Equation 31}$$

$$P + B \underset{K_4}{\overset{K_3}{\rightleftharpoons}} PB \qquad \textbf{Equation 32}$$

We make the assumption that A does not form a complex with B and that the enzyme does not chemically modify either ligand.

The rate equations for the formation of the complex can be described by equations 33 and 34.

$$\frac{d\ PA}{dt} = k_1\ P\ \ A\ - k_2\ PA \qquad \textbf{Equation 33}$$

$$\frac{d\ PB}{dt} = k_3\ P\ \ B\ - k_4\ PB \qquad \textbf{Equation 34}$$

Where: [P] is the free concentration of the protein; [A] the concentration of free ligand A; [B] the concentration of free ligand B; [PA] the concentration of the complex of A with P; [PB] the concentration of the complex of B with P; K_1 the on-rate for ligand A; K_2 the off-rate for ligand A; K_3 the on-rate for ligand B and K_4 the off-rate for ligand B.

Equations of this form can be solved using the differential equation solving program Berkeley Madonna (V.8.3.11, Macy and Oster). The program enables the user to model the time course of the reaction if the initial concentrations of the protein (P_{total}) and ligands are entered (A_{total} and B_{total}) and the forward (on-rates, k_1 and k_3) and reverse (off-rates, k_2 and

k_4) rate constants are known. These parameters are entered with the equations describing the equilibria between the species (equations 31 and 32 in the above example) and the time course for the simulation. This study used the chemical reactions module of program; this module automatically generates the rate equations. The on and off-rates have been published for the interaction between FKBP12 and rapamycin; K_{on} = 0.75 ± 0.04 $\mu M^{-1}s^{-1}$ and K_{off} = 1.6 ± 0.2 ($x10^{-3}s^{-1}$) (Wear and Walkinshaw, 2007). Rapamycin is a large (MW = 914 Da), high affinity ligand (K_d = 0.2 nM). These results were for the same construct of 6H-FKBP12 used in this study by surface plasmon resonance. A similar experiment was carried out for the much smaller, lower affinity ligand SD1 (6.4.5, Figure 6-10 SPR FKBP12-SD1). This compound exhibited an on-and off-rate that were too fast to quantify (> 10 $\mu M^{-1}s^{-1}$). On-rates are typically in the range 10^3 to 10^6 $M^{-1}s^{-1}$ and off-rates 10^{-1} to 5 $x10^{-6}$ s^- for protein ligand interactions. Unless the on and off-rates are accurately defined in the program it is impossible to accurately calculate the concentration of a given species before equilibrium is attained. This information was not known for the simulations carried out for the majority of the ligands and as ligands were not part of a chemically similar series of known kinetics. The decision was made to use the simulation to calculate the relative proportions of the species in solution at equilibrium, when the ratio of the off-rate/on-rate (K_d) is important rather than the specific values. The on-rate entered was entered was, 10^3 $\mu M^{-1}s^{-1}$, close to the diffusion rate limit (Scaronolc k. and Stockmayer W.H., 1973). The off-rate calculated from equation 35 using the K_d of the ligand in question.

$$K_d = \frac{k_{off}}{k_{on}} \qquad \textbf{Equation 35}$$

11 Appendix 2 – The fragment library

<Fragment number>_<Maybridge, UK catalogue number>

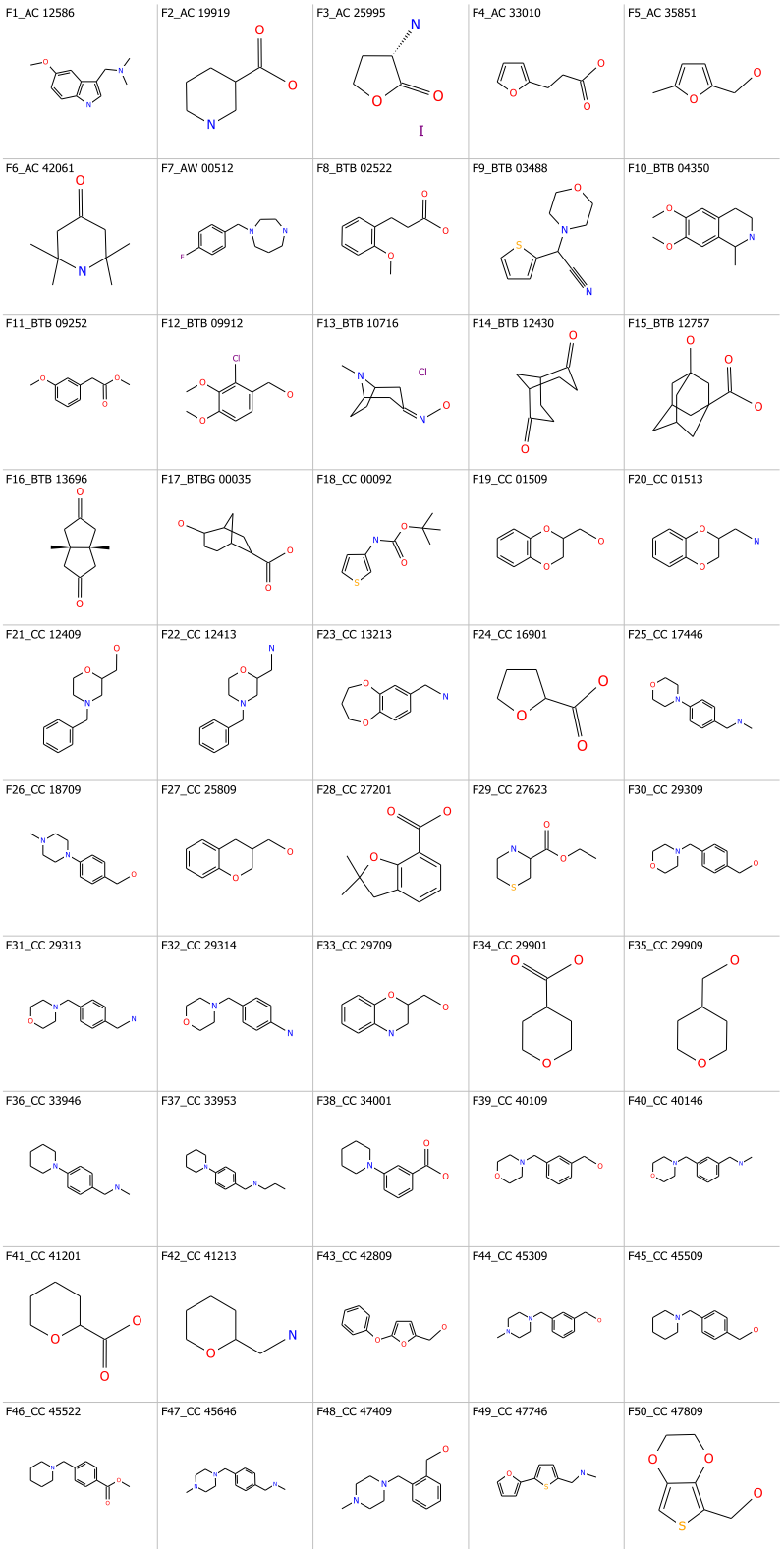


Figure 11-1 Fragment library (F1-F50)

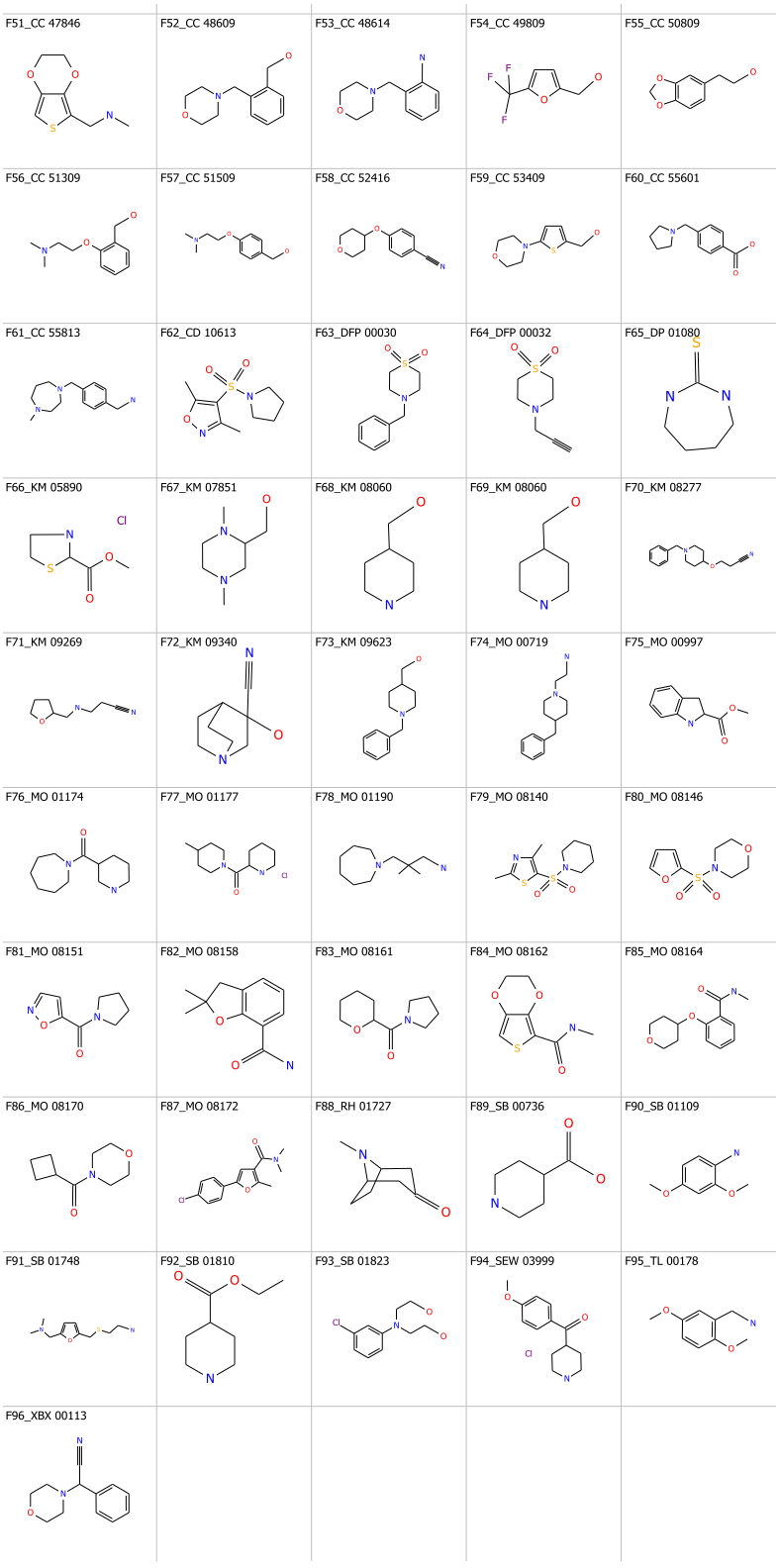
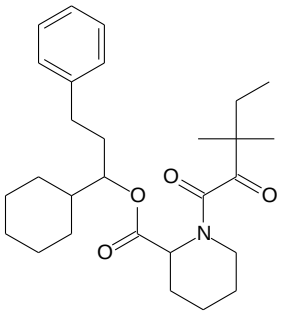
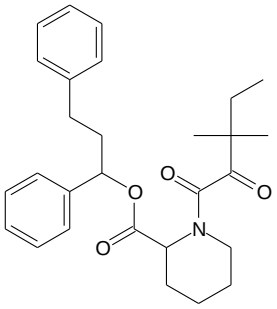
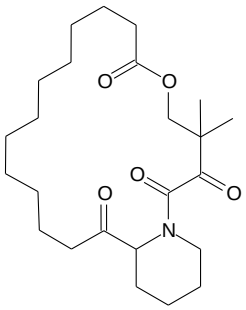
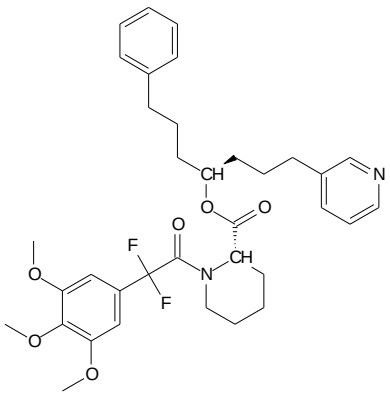
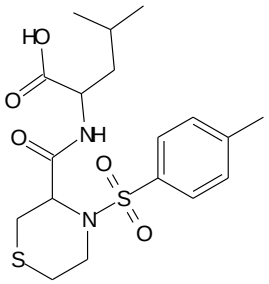
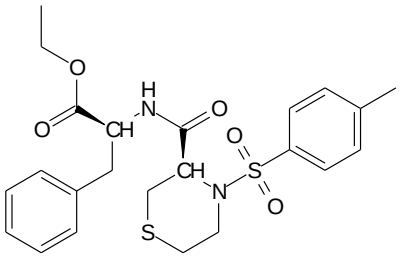
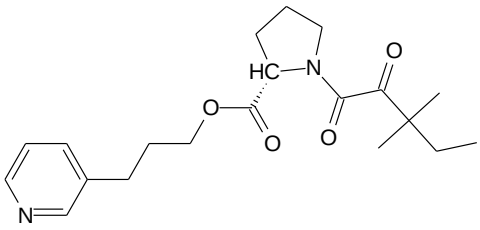
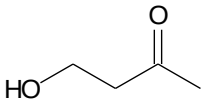
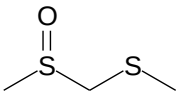
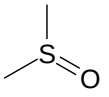


Figure 11-2 Fragment library (F51-F96)

12 Appendix 3 – Inhibitors crystallised in complex with FKBP12

PDB	*Structure*	*Ligand (reference)*	$K_{i,app}$ *(μM)*
2DG3	**Figure 1-7**	rapamycin (Fulton *et al.*, 2003)	0.0002
1fkf	**Figure 1-7**	FK506 (Van Duyne *et al.*, 1993)	0.0004
1fkh		SBX (Holt *et al.*, 1993)	0.3 – 0.6
1fkg		SB3 (Holt *et al.*, 1993)	0.3 ± 0.04
1fki		SBI (Holt *et al.*, 1993)	0.1± 0.02
1j4r		FK001(Dubowchik *et al.*, 2001)	0.019

PDB	Structure	Ligand (reference)	$K_{i,app}$ (μM)
1j4i		308 (Sun *et al.*, 2003)	
1j4h		107 (Sun *et al.*, 2003)	
1f40		GPI1046 (Sich *et al.*, 2000)	
1d7j		4-hydroxy-butanone (Burkhard *et al.*, 2000)	
1d7i		DSS (Burkhard *et al.*, 2000)	
1d7h		DMSO (Burkhard *et al.*, 2000)	20000 (37000±14000)

Reference List

Abraham RT, Wiederrecht GJ (1996). Immunopharmacology of rapamycin. *Annu Rev Immunol* **14:** 483-510.

Achim CL, Masliah E, Schindelar J, Avramut M (2004). Immunophilin expression in the HIV-infected brain. *J Neuroimmunol* **157:** 126-132.

Angelopoulos N, Hadjiprocopis A, Walkinshaw MD (2009). Baysian Model Averaging for Ligand Discovery. *Journal of Chemical Information and Modeling* ***in press***.

Armistead DM, Badia MC, Deininger DD, Duffy JP, Saunders JO, Tung RD *et al.* (1995). Design, synthesis and structure of non-macrocyclic inhibitors of FKBP12, the major binding protein for the immunosuppressant FK506. *Acta Crystallogr D Biol Crystallogr* **51:** 522-528.

Atkins PW. (1986). Physical Chemistry. Oxford University Press.

Babine RE, Bender SL (1997). Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design. *Chem Rev* **97:** 1359-1472.

Bai X, Ma D, Liu A, Shen X, Wang QJ, Liu Y *et al.* (2007). Rheb activates mTOR by antagonizing its endogenous inhibitor, FKBP38. *Science* **318:** 977-980.

Bajorath J (2002). Integration of virtual and high-throughput screening. *Nat Rev Drug Discov* **1:** 882-894.

Ballester PJ, Richards WG (2007). Ultrafast shape recognition to search compound databases for similar molecular shapes. *J Comput Chem* **28:** 1711-1723.

Banerjee A, Periyasamy S, Wolf IM, Hinds TD, Jr., Yong W, Shou W *et al.* (2008). Control of glucocorticoid and progesterone receptor subcellular localization by the ligand-binding domain is mediated by distinct interactions with tetratricopeptide repeat proteins. *Biochemistry* **47:** 10471-10480.

Barik S (2006). Immunophilins: for the love of proteins. *Cell Mol Life Sci* **63:** 2889-2900.

Barnard JM, Downs GM (1997). Chemical Fragment Generation and Clustering Softwares. *J Chem Inf Comput Sci* **37:** 141-142.

Bauer S, Shiloach J (1974). Maximal exponential growth rate and yield of *E. coli* obtainable in a bench-scale fermentor. *Biotechnol Bioeng* **16:** 933-941.

Bell A, Monaghan P, Page AP (2006). Peptidyl-prolyl cis-trans isomerases (immunophilins) and their roles in parasite biochemistry, host-parasite interaction and antiparasitic drug action. *Int J Parasitol* **36:** 261-276.

Bender A, Glen RC (2004). Molecular similarity: a key technique in molecular informatics. *Org Biomol Chem* **2:** 3204-3218.

Benkestock K, Sundqvist G, Edlund PO, Roeraade J (2004). Influence of droplet size, capillary-cone distance and selected instrumental parameters for the analysis of noncovalent protein-ligand complexes by nano-electrospray ionization mass spectrometry. *J Mass Spectrom* **39:** 1059-1067.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res* **28:** 235-242.

Bierer BE, Mattila PS, Standaert RF, Herzenberg LA, Burakhoff SJ, Crabtree G *et al.* (1990). Two Distinct Signal Transmission Pathways in T Lymphocytes are Inhibited by Complexes Formed between an Immunophilin and either FK506 or Rapamycin. *Proceedings of the National Academy of Sciences* **87:** 9231-9235.

Blow DM, Birktoft JJ, Hartley BS (1969). Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* **221:** 337-340.

Boehm HJ (1998). Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J Comput Aided Mol Des* **12:** 309-323.

Boehm HJ, Flohr A, Stahl M (2004). Scaffold hopping. *Drug Discovery Today: Technologies* **1:** 217-224.

Bondi A (1964). van der Waals Volumes and Radii. *J Phys Chem* **68:** 441-451.

Bouchemal K (2008). New challenges for pharmaceutical formulations and drug delivery systems characterization using isothermal titration calorimetry. *Drug Discov Today*.

Brandts JF, Lin LN (1990). Study of strong to ultratight protein interactions using differential scanning calorimetry. *Biochemistry (Mosc)* **29:** 6927-6940.

Braun W, Kallen J, Mikol V, Walkinshaw MD, Wuthrich K (1995). Three-dimensional structure and actions of immunosuppressants and their immunophilins. *FASEB J* **9:** 63-72.

Brecht S, Schwarze K, Waetzig V, Christner C, Heiland S, Fischer G *et al.* (2003). Changes in peptidyl-prolyl cis/trans isomerase activity and FK506 binding protein expression following neuroprotection by FK506 in the ischemic rat brain. *Neuroscience* **120:** 1037-1048.

Brooijmans N, Kuntz ID (2003). Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **32:** 335-373.

Brown RE, Jarvis KL, Hyland KJ (1989). Protein measurement using bicinchoninic acid: elimination of interfering substances. *Anal Biochem* **180:** 136-139.

Brown SP, Hajduk PJ (2006). Effects of conformational dynamics on predicted protein druggability. *ChemMedChem* **1:** 70-72.

Burkhard P, Hommel U, Sanner M, Walkinshaw MD (1999). The discovery of steroids and other novel FKBP inhibitors using a molecular docking program. *J Mol Biol* **287:** 853-858.

Burkhard P, Taylor P, Walkinshaw MD (2000). X-ray structures of small ligand-FKBP complexes provide an estimate for hydrophobic interaction energies. *J Mol Biol* **295:** 953-962.

C.Elegans sequencing consortium (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* **282:** 2012-2018.

Cabani S, Gianni P, Mollica V, Lepori L (1981). Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *J Solution Chem* **10:** 563-595.

Carbonell T, Freire E (2005). Binding thermodynamics of statins to HMG-CoA reductase. *Biochemistry (Mosc)* **44:** 11741-11748.

Cardamone M, Puri NK (1992). Spectrofluorimetric assessment of the surface hydrophobicity of proteins. *Biochem J* **282 (Pt 2):** 589-593.

Carr RA, Congreve M, Murray CW, Rees DC (2005). Fragment-based lead discovery: leads by design. *Drug Discov Today* **10:** 987-992.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50:** 760-763.

Cerny J, Hobza P (2007). Non-covalent interactions in biomacromolecules. *Phys Chem Chem Phys* **9:** 5291-5303.

Chadli A, Bruinsma ES, Stensgard B, Toft D (2008). Analysis of Hsp90 cochaperone interactions reveals a novel mechanism for TPR protein recognition. *Biochemistry* **47:** 2850-2857.

Chaires JB (1997). Possible origin of differences between van't Hoff and calorimetric enthalpy estimates. *Biophys Chem* **64:** 15-23.

Chang CE, Chen W, Gilson MK (2007). Ligand configurational entropy and protein binding. *Proc Natl Acad Sci U S A* **104:** 1534-1539.

Chelu MG, Danila CI, Gilman CP, Hamilton SL (2004). Regulation of ryanodine receptors by FK506 binding proteins. *Trends Cardiovasc Med* **14:** 227-234.

Cheng T, Zhao Y, Li X, Lin F, Xu Y, Zhang X *et al.* (2007). Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J Chem Inf Model* **47:** 2140-2148.

Cheng Y, Prusoff WH (1973). Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem Pharmacol* **22:** 3099-3108.

Choi C, Li JH, Vaal M, Thomas C, Limburg D, Wu YQ *et al.* (2002). Use of parallel-synthesis combinatorial libraries for rapid identification of potent FKBP12 inhibitors. *Bioorg Med Chem Lett* **12:** 1421-1428.

Christner C, Herdegen T, Fischer G (2001). FKBP ligands as novel therapeutics for neurological disorders. *Mini Rev Med Chem* **1:** 377-397.

Christner C, Wyrwa R, Marsch S, Kullertz G, Thiericke R, Grabley S *et al.* (1999). Synthesis and cytotoxic evaluation of cycloheximide derivatives as potential inhibitors of FKBP12 with neuroregenerative properties. *J Med Chem* **42:** 3615-3622.

Cimmperman P, Baranauskiene L, Jachimoviciute S, Jachno J, Torresan J, Michailoviene V *et al.* (2008). A quantitative model of thermal stabilization and destabilization of proteins by ligands. *Biophys J* **95:** 3222-3231.

Ciulli A, Abell C (2007). Fragment-based approaches to enzyme inhibition. *Curr Opin Biotechnol* **18:** 489-496.

Congreve M, Chessari G, Tisi D, Woodhead AJ (2008). Recent developments in fragment-based drug discovery. *J Med Chem* **51:** 3661-3680.

Connelly PR, Thomson JA (1992). Heat capacity changes and hydrophobic interactions in the binding of FK506 and rapamycin to the FK506 binding protein. *Proc Natl Acad Sci U S A* **89:** 4781-4785.

Connelly PR, Thomson JA, Fitzgibbon MJ, Bruzzese FJ (1993). Probing hydration contributions to the thermodynamics of ligand binding by proteins. Enthalpy and heat capacity changes of tacrolimus and rapamycin binding to FK506 binding protein in D2O and H2O. *Biochemistry (Mosc)* **32:** 5583-5590.

Cooper A (1999a). Thermodynamic analysis of biomolecular interactions. *Curr Opin Chem Biol* **3:** 557-563.

Cooper A (1999b). Thermodynamics of protein folding and Stability.In *Protein: A Comprehensive Treatise.* pp. 217-270. JAI Press Inc.

Cooper A (2005). Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics. *Biophys Chem* **115:** 89-97.

Cooper A, Cameron D, Jakus J, Pettigrew GW (2007). Pressure perturbation calorimetry, heat capacity and the role of water in protein stability and interactions. *Biochem Soc Trans* **35:** 1547-1550.

Cooper A, Johnson CM (1994). Introduction to microcalorimetry and biomolecular energetics. *Methods Mol Biol* **22:** 109-124.

Cooper A, Johnson CM, Lakey JH, Nollmann M (2001). Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys Chem* **93:** 215-230.

Cooper MA (2002). Optical biosensors in drug discovery. *Nat Rev Drug Discov* **1:** 515-528.

Crum Brown A, Fraser TR (1869). On the Connection between Chemical Constitution and Physiological Action. Part I. - On the Physiological Action of the Salts of the Ammonium Bases, derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. *Trans Roy Soc Edinburgh***:** 151-203.

Daniel JM, Friess SD, Rajagopalan S, Wendt S, Zenobi R (2002). Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry. *International Journal of Mass Spectrometry* **216:** 1-27.

Davies TH, Sanchez ER (2005). FKBP52. *Int J Biochem Cell Biol* **37:** 42-47.

Dearden JC (2003). In silico prediction of drug toxicity. *J Comput Aided Mol Des* **17:** 119-127.

Deivanayagam CC, Carson M, Thotakura A, Narayana SV, Chodavarapu RS (2000). Structure of FKBP12.6 in complex with rapamycin. *Acta Crystallogr D Biol Crystallogr* **56:** 266-271.

Delaney JS (2005). Predicting aqueous solubility from structure. *Drug Discov Today* **10:** 289-295.

Diller DJ (2008). The synergy between combinatorial chemistry and high-throughput screening. *Curr Opin Drug Discov Devel* **11:** 346-355.

Dole M, Mack LL, Hines RL, Mobley RC, Ferguson LD, Alice MB (1968). Molecular Beams of Macroions. *The Journal of Chemical Physics* **49:** 2240-2249.

Dornan J, Taylor P, Walkinshaw MD (2003). Structures of immunophilins and their ligand complexes. *Curr Top Med Chem* **3:** 1392-1409.

Dragovich PS, Barker JE, French J, Imbacuan M, Kalish VJ, Kissinger CR *et al.* (1996). Structure-based design of novel, urea-containing FKBP12 inhibitors. *J Med Chem* **39:** 1872-1884.

Driggers EM, Hale SP, Lee J, Terrett NK (2008). The exploration of macrocycles for drug discovery--an underexploited structural class. *Nat Rev Drug Discov* **7:** 608-624.

Dubowchik GM, Vrudhula VM, Dasgupta B, Ditta J, Chen T, Sheriff S *et al.* (2001). 2-Aryl-2,2-difluoroacetamide FKBP12 ligands: synthesis and X-ray structural studies. *Org Lett* **3:** 3987-3990.

Durant JL, Leland BA, Henry DR, Nourse JG (2002). Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* **42:** 1273-1280.

Eckert H, Bajorath J (2007). Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov Today* **12:** 225-233.

Edlich F, Weiwad M, Wildemann D, Jarczowski F, Kilka S, Moutty MC *et al.* (2006). The specific FKBP38 inhibitor N-(N',N'-dimethylcarboxamidomethyl)cycloheximide has potent neuroprotective and neurotrophic properties in brain ischemia. *J Biol Chem* **281:** 14961-14970.

Egan DA, Logan TM, Liang H, Matayoshi E, Fesik SW, Holzman TF (1993). Equilibrium denaturation of recombinant human FK binding protein in urea. *Biochemistry (Mosc)* **32:** 1920-1927.

Emsley P, Cowtan K (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126-2132.

Epps DE, Sarver RW, Rogers JM, Herberg JT, Tomich PK (2001). The ligand affinity of proteins measured by isothermal denaturation kinetics. *Anal Biochem* **292:** 40-50.

Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357:** 289-298.

Ertl P, Jelfs S, Muhlbacher J, Schuffenhauer A, Selzer P (2006). Quest for the rings. In silico exploration of ring universe to identify novel bioactive heteroaromatic scaffolds. *J Med Chem* **49:** 4568-4573.

Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* **15:** 411-428.

Feher M (2006). Consensus scoring for protein-ligand interactions. *Drug Discov Today* **11:** 421-428.

Felitsyn N, Peschke M, Kebarle P (2002). Origin and number of charges observed on multiply-protonated native proteins produced by ESI. *International Journal of Mass Spectrometry* **219:** 39-62.

Ferguson WJ, Braunschweiger KI, Braunschweiger WR, Smith JR, McCormick JJ, Wasmann CC *et al.* (1980). Hydrogen ion buffers for biological research. *Anal Biochem* **104:** 300-310.

Fernandez de la Mora J (2000). Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal Chim Acta* **406:** 93-104.

Fischer G (2000). Chemical aspects of peptide bond isomerisation. *Chem Soc Rev* **29:** 119-127.

Fischer G, Bang H, Berger E, Schellenberger A (1984a). Conformational specificity of chymotrypsin toward proline-containing substrates. *Biochim Biophys Acta* **791:** 87-97.

Fischer G, Bang H, Mech C (1984b). Determination of enzymatic catalysis for the cis-trans-isomerization of peptide binding in proline-containing peptides. *Biomed Biochim Acta* **43:** 1101-1111.

Fischer S, Michnick S, Karplus M (1993). A mechanism for rotamase catalysis by the FK506 binding protein (FKBP). *Biochemistry (Mosc)* **32:** 13830-13837.

Freire E (2008). Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov Today* **19:** 869-874.

Freyer MW, Lewis EA (2008). Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol* **84:** 79-113.

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT *et al.* (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47:** 1739-1749.

Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA *et al.* (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **49:** 6177-6196.

Frottin F, Martinez A, Peynot P, Mitra S, Holz RC, Giglione C *et al.* (2006). The proteomics of N-terminal methionine cleavage. *Mol Cell Proteomics* **5:** 2336-2349.

Fulton KF, Jackson SE, Buckle AM (2003). Energetic and structural analysis of the role of tryptophan 59 in FKBP12. *Biochemistry (Mosc)* **42:** 2364-2372.

Fulton KF, Main ER, Daggett V, Jackson SE (1999). Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J Mol Biol* **291:** 445-461.

Gabelica V, De PE (2005). Internal energy and fragmentation of ions produced in electrospray sources. *Mass Spectrom Rev* **24:** 566-587.

Galat A (2003). Peptidylprolyl cis/trans isomerases (immunophilins): biological diversity--targets--functions. *Curr Top Med Chem* **3:** 1315-1347.

Galat A (2008). Functional drift of sequence attributes in the FK506-binding proteins (FKBPs). *J Chem Inf Model* **48:** 1118-1130.

Galat A, Lane WS, Standaert RF, Schreiber SL (1992). A rapamycin-selective 25-kDa immunophilin. *Biochemistry (Mosc)* **31:** 2427-2434.

Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI *et al.* (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87:** 1285-1294.

Ganen B, Li Y-T, Henion J (1991). Detection of Noncovalent Receptor-Ligand Complexes by Mass Spectrometry. *J Am Chem Soc* **113:** 6294-6296.

Ganesan A (2008). The impact of natural products upon modern drug discovery. *Curr Opin Chem Biol* **12:** 306-317.

Gao X, Hu H (2008). Quality control of the proteins associated with neurodegenerative diseases. *Acta Biochim Biophys Sin (Shanghai)* **40:** 612-618.

Garcia-Hernandez E, Zubillaga RA, Chavelas-Adame EA, Vazquez-Contreras E, Rojo-Dominguez A, Costas M (2003). Structural energetics of protein-carbohydrate interactions: Insights derived from the study of lysozyme binding to its natural saccharide inhibitors. *Protein Sci* **12:** 135-142.

Gasteiger E, Hoogland C, Gattiker A, Duvald S, Wilkins M, Appel R & Barioch A (2005). Protein Identification and Analysis Tools on the ExPASy Server. pp. 571-607. Humana Press.

Gething MJ, Sambrook J (1992). Protein folding in the cell. *Nature* **355:** 33-45.

Goodford PJ (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **28:** 849-857.

Greene J, Kahn S, Savoj H, Sprague P, Teig S (1994). Chemical Function Queries for 3D Database Search. *J Chem Inf Comput Sci* **34:** 1297-1308.

Hacker J, Fischer G (1993). Immunophilins: structure-function relationship and possible role in microbial pathogenicity. *Mol Microbiol* **10:** 445-456.

Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT *et al.* (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **47:** 1750-1759.

Hamelberg D, McCammon JA (2009). Mechanistic insight into the role of transition-state stabilization in cyclophilin A. *J Am Chem Soc* **131:** 147-152.

Hamilton GS, Steiner JP (1998). Immunophilins: beyond immunosuppression. *J Med Chem* **41:** 5119-5143.

Handschumacher RE, Harding MW, Rice J, Drugge RJ, Speicher DW (1984). Cyclophilin: a specific cytosolic binding protein for cyclosporin A. *Science* **226:** 544-547.

Hann MM, Oprea TI (2004). Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol* **8:** 255-263.

Haranczyk M, Holliday J (2008). Comparison of similarity coefficients for clustering and compound selection. *J Chem Inf Model* **48:** 498-508.

Harding MW, Galat A, Uehling DE, Schreiber SL (1989). A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* **341:** 758-760.

Harrison RK, Stein RL (1990). Substrate specificities of the peptidyl prolyl cis-trans isomerase activities of cyclophilin and FK-506 binding protein: evidence for the existence of a family of distinct enzymes. *Biochemistry (Mosc)* **29:** 3813-3816.

Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H (2005). Fragment-based lead discovery using X-ray crystallography. *J Med Chem* **48:** 403-413.

Hawe A, Sutter M, Jiskoot W (2008). Extrinsic fluorescent dyes as tools for protein characterization. *Pharm Res* **25:** 1487-1499.

Helms V, Wade RC (1998). Hydration energy landscape of the active site cavity in cytochrome P450cam. *Proteins* **32:** 381-396.

Hernandez H, Robinson CV (2007). Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat Protoc* **2:** 715-726.

Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E *et al.* (2006). New methods for ligand-based virtual screening: use of data fusion and machine learning to enhance the effectiveness of similarity searching. *J Chem Inf Model* **46:** 462-470.

Hesterkamp T, Whittaker M (2008). Fragment-based activity space: smaller is better. *Curr Opin Chem Biol* **12:** 260-268.

Himukai R, Kuzuhara T, Horikoshi M (1999b). Relationship between the subcellular localization and structures of catalytic domains of FKBP-type PPIases. *J Biochem* **126:** 879-888.

Himukai R, Kuzuhara T, Horikoshi M (1999a). Relationship between the subcellular localization and structures of catalytic domains of FKBP-type PPIases. *J Biochem* **126:** 879-888.

Hintersteiner M, Auer M (2008). Single-bead, single-molecule, single-cell fluorescence: technologies for drug screening and target validation. *Ann N Y Acad Sci* **1130:** 1-11.

Hinton AC. (2005). Database Mining: EDULISS a descriptor based approach, Edinburgh University Thesis.

Hofstadler SA, Sannes-Lowery KA (2006). Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes. *Nat Rev Drug Discov* **5:** 585-595.

Holdgate GA (2001). Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. *BioTechniques* **31:** 164-6, 168, 170.

Holt DA, Konialian-Beck AL, Oh HJ, Yen HK, Rozamus LW, Krog AJ *et al.* (1994). Structure-activity studies of synthetic FKBP ligands as peptidyl-prolyl isomerase inhibitors. *Bioorg Med Chem Lett* **4:** 315-320.

Holt DA, Luengo JI, Yamashita DS, Oh HJ, Konialian AL, Yen HK *et al.* (1993). Design, synthesis, and kinetic evaluation of high-affinity FKBP ligands and the X-ray crystal structures of their complexes with FKBP12. *J Am Chem Soc* **115:** 9925-9938.

Homans SW (2007). Water, water everywhere--except where it matters? *Drug Discov Today* **12:** 534-539.

Hsin K. (2009). EDULISS, Edinburgh University Thesis.

Hudack RA, Jr., Barta NS, Guo C, Deal J, Dong L, Fay LK *et al.* (2006). Design, synthesis, and biological activity of novel polycyclic aza-amide FKBP12 ligands. *J Med Chem* **49:** 1202-1206.

Iribarne JV, Thomson BA (1976). On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics* **64:** 2287-2294.

Irwin JJ, Shoichet BK (2005). ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **45:** 177-182.

Isaacs ED, Shukla A, Platzman PM, Hamann DR, Barbiellini B, Tulk CA (1999). Covalency of the Hydrogen Bond in Ice: A Direct X-Ray Measurement. *Phys Rev Lett* **82:** 600.

Ishikawa K, Nagase T, Suyama M, Miyajima N, Tanaka A, Kotani H *et al.* (1998). Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. *DNA Res* **5:** 169-176.

Ivery MT (2000). Immunophilins: switched on protein binding domains? *Med Res Rev* **20:** 452-484.

Jarczowski F, Fischer G, Edlich F (2008a). FKBP36 forms complexes with clathrin and Hsp72 in spermatocytes. *Biochemistry* **47:** 6946-6952.

Jarczowski F, Jahreis G, Erdmann F, Schierhorn A, Fischer G, Edlich F (2008b). FKBP36 is an inherent multi-functional glyceraldehyde-3-phosphate dehydrogenase inhibitor. *J Biol Chem*.

Jecklin MC, Touboul D, Bovet C, Wortmann A, Zenobi R (2008). Which electrospray-based ionization method best reflects protein-ligand interactions found in solution? a comparison of ESI, nanoESI, and ESSI for the determination of dissociation constants with mass spectrometry. *J Am Soc Mass Spectrom* **19:** 332-343.

Jelesarov I, Bosshard HR (1999). Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J Mol Recognit* **12:** 3-18.

Johansen MB, Kiemer L, Brunak S (2006). Analysis and prediction of mammalian protein glycation. *Glycobiology* **16:** 844-853.

Jones G, Willett P, Glen RC (1995). A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J Comput Aided Mol Des* **9:** 532-549.

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997). Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267:** 727-748.

Jorgensen WL, Duffy EM (2002). Prediction of drug solubility from structure. *Adv Drug Deliv Rev* **54:** 355-366.

Kallen J, Mikol V, Taylor P, Walkinshaw MD (1998). X-ray structures and analysis of 11 cyclosporin derivatives complexed with cyclophilin A. *J Mol Biol* **283:** 435-449.

Kang CB, Hong Y, Dhe-Paganon S, Yoon HS (2008). FKBP family proteins: immunophilins with versatile biological functions. *Neurosignals* **16:** 318-325.

Karp DA, Gittis AG, Stahley MR, Fitch CA, Stites WE, Garcia-Moreno EB (2007). High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp. *Biophys J* **92:** 2041-2053.

Ke HM, Zydowsky LD, Liu J, Walsh CT (1991). Crystal structure of recombinant human T-cell cyclophilin A at 2.5 A resolution. *Proc Natl Acad Sci U S A* **88:** 9483-9487.

Kitchen DB, Decornez H, Furr JR, Bajorath J (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **3:** 935-949.

Klebe G, Boehm HJ (1997). Energetic and entropic factors determining binding affinity in protein-ligand complexes. *J Recept Signal Transduct Res* **17:** 459-473.

Kofron JL, Kuzmic P, Kishore V, Colon-Bonilla E, Rich DH (1991). Determination of kinetic constants for peptidyl prolyl cis-trans isomerases by an improved spectrophotometric assay. *Biochemistry (Mosc)* **30:** 6127-6134.

Konstantinou-Kirtay C, Mitchell JB, Lumley JA (2007). Scoring functions and enrichment: a case study on Hsp90. *BMC Bioinformatics* **8:** 27.

Kontopidis G, Taylor P, Walkinshaw MD (2004). Enzymatic and structural characterization of non-peptide ligand-cyclophilin complexes. *Acta Crystallogr D Biol Crystallogr* **60:** 479-485.

Krishnamurthy VM, Kaufman GK, Urbach AR, Gitlin I, Gudiksen KL, Weibel DB *et al.* (2008). Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem Rev* **108:** 946-1051.

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982). A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **161:** 269-288.

Kuntz ID, Chen K, Sharp KA, Kollman PA (1999). The maximal affinity of ligands. *Proc Natl Acad Sci U S A* **96:** 9997-10002.

Kuzhandaivelu N, Cong YS, Inouye C, Yang WM, Seto E (1996). XAP2, a novel hepatitis B virus X-associated protein that inhibits X transactivation. *Nucleic Acids Res* **24:** 4741-4750.

Laemmli U (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227:** 680-685.

Lafont V, Armstrong AA, Ohtaka H, Kiso Y, Mario AL, Freire E (2007). Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chem Biol Drug Des* **69:** 413-422.

Lam E, Martin MM, Timerman AP, Sabers C, Fleischer S, Lukas T *et al.* (1995). A novel FK506 binding protein can mediate the immunosuppressive effects of FK506 and is associated with the cardiac ryanodine receptor. *J Biol Chem* **270:** 26511-26522.

Laurence C, Brameld KA, Graton J, Le Questel JY, Renault E (2009). The pK(BHX) database: toward a better understanding of hydrogen-bond basicity for medicinal chemists. *J Med Chem* **52:** 4073-4086.

Lebedev AA, Vagin AA, Murshudov GN (2008). Model preparation in MOLREP and examples of model improvement using X-ray data. *Acta Crystallogr D Biol Crystallogr* **64:** 33-39.

Li L, Dantzer JJ, Nowacki J, O'Callaghan BJ, Meroueh SO (2008). PDBcal: a comprehensive dataset for receptor-ligand interactions with three-dimensional structures and binding thermodynamics from isothermal titration calorimetry. *Chem Biol Drug Des* **71:** 529-532.

Liang J, Hung DT, Schreiber SL, Clardy J (1996). Structure of the Human 25 kDa FK506 Binding Protein Complexed with Rapamycin. *Journal of the American Chemical Society* **118:** 1231-1232.

Limburg DC, Thomas BE, Li JH, Fuller M, Spicer D, Chen Y *et al.* (2003). Synthesis and evaluation of chiral bicyclic proline FKBP12 ligands. *Bioorg Med Chem Lett* **13:** 3867-3870.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Delivery Rev* **23:** 3-25.

Lisgarten JN, Gupta V, Maes D, Wyns L, Zegers I, Palmer RA *et al.* (1993). Structure of the crystalline complex of cytidylic acid (2'-CMP) with ribonuclease at 1.6 A resolution. Conservation of solvent sites in RNase-A high-resolution structures. *Acta Crystallogr D Biol Crystallogr* **49:** 541-547.

Lo MC, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M *et al.* (2004). Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* **332:** 153-159.

Loo JA (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom Rev* **16:** 1-23.

Lu KP, Finn G, Lee TH, Nicholson LK (2007). Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol* **3:** 619-629.

Lyons WE, George EB, Dawson TM, Steiner JP, Snyder SH (1994). Immunosuppressant FK506 promotes neurite outgrowth in cultures of PC12 cells and sensory ganglia. *Proc Natl Acad Sci U S A* **91:** 3191-3195.

Main ER, Fulton KF, Jackson SE (1999). Folding pathway of FKBP12 and characterisation of the transition state. *J Mol Biol* **291:** 429-444.

Matulis D, Kranz JK, Salemme FR, Todd MJ (2005). Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry (Mosc)* **44:** 5258-5266.

McInnes C, Mazumdar A, Mezna M, Meades C, Midgley C, Scaerou F *et al.* (2006). Inhibitors of Polo-like kinase reveal roles in spindle-pole maintenance. *Nat Chem Biol* **2:** 608-617.

Moriguchi I., Hirono S., Liu Q., Nakagome Y., Matsushita Y. (1992). Simple methods of calculating octanol water partition coefficient. *Chem Pharm Bull* **40:** 127-130.

Morrison JF (1969). Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim Biophys Acta* **185:** 269-286.

Munoz V, Sanchez-Ruiz JM (2004). Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. *Proc Natl Acad Sci U S A* **101:** 17646-17651.

Murshudov GN, Vagin AA, Dodson EJ (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240-255.

Naghibi H, Tamura A, Sturtevant JM (1995). Significant discrepancies between van't Hoff and calorimetric enthalpies. *Proc Natl Acad Sci U S A* **92:** 5597-5599.

Nair SC, Rimerman RA, Toran EJ, Chen S, Prapapanich V, Butts RN *et al.* (1997). Molecular cloning of human FKBP51 and comparisons of immunophilin interactions with Hsp90 and progesterone receptor. *Mol Cell Biol* **17:** 594-603.

Nakamura T, Yabe D, Kanazawa N, Tashiro K, Sasayama S, Honjo T (1998). Molecular cloning, characterization, and chromosomal localization of FKBP23, a novel FK506-binding protein with Ca2+-binding ability. *Genomics* **54:** 89-98.

Navratilova I, Papalia GA, Rich RL, Bedinger D, Brophy S, Condon B *et al.* (2007). Thermodynamic benchmark study using Biacore technology. *Anal Biochem* **364:** 67-77.

Neumann T, Junker HD, Schmidt K, Sekul R (2007). SPR-based fragment screening: advantages and applications. *Curr Top Med Chem* **7:** 1630-1642.

Neye H, Verspohl EJ (2004). The FK506 binding protein 13 kDa (FKBP13) interacts with the C-chain of complement C1q. *BMC Pharmacol* **4:** 19.

Nigam SK, Jin YJ, Jin MJ, Bush KT, Bierer BE, Burakoff SJ (1993). Localization of the FK506-binding protein, FKBP 13, to the lumen of the endoplasmic reticulum. *Biochem J* **294 (Pt 2):** 511-515.

Norinder U, Bergstrom CA (2006). Prediction of ADMET Properties. *ChemMedChem* **1:** 920-937.

Oprea TI, Allu TK, Fara DC, Rad RF, Ostopovici L, Bologa CG (2007). Lead-like, drug-like or "Pub-like": how different are they? *J Comput Aided Mol Des* **21:** 113-119.

Oprea TI, Matter H (2004). Integrating virtual screening in lead discovery. *Curr Opin Chem Biol* **8:** 349-358.

Page AP, Kumar S, Carlow CK (1995). Parasite cyclophilins and antiparasite activity of cyclosporin A. *Parasitol Today* **11:** 385-388.

Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E *et al.* (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **6:** 429-440.

Park ST, Aldape RA, Futer O, DeCenzo MT, Livingston DJ (1992). PPIase catalysis by human FK506-binding protein proceeds through a conformational twist mechanism. *J Biol Chem* **267:** 3316-3324.

Patriksson A, Marklund E, van der SD (2007). Protein structures under electrospray conditions. *Biochemistry (Mosc)* **46:** 933-945.

Patterson CE, Abrams WR, Wolter NE, Rosenbloom J, Davis EC (2005). Developmental regulation and coordinate reexpression of FKBP65 with extracellular matrix proteins after lung injury suggest a specialized function for this endoplasmic reticulum immunophilin. *Cell Stress Chaperones* **10:** 285-295.

Patterson CE, Gao J, Rooney AP, Davis EC (2002a). Genomic organization of mouse and human 65 kDa FK506-binding protein genes and evolution of the FKBP multigene family. *Genomics* **79:** 881-889.

Patterson CE, Gao J, Rooney AP, Davis EC (2002b). Genomic organization of mouse and human 65 kDa FK506-binding protein genes and evolution of the FKBP multigene family. *Genomics* **79:** 881-889.

Peattie DA, Harding MW, Fleming MA, DeCenzo MT, Lippke JA, Livingston DJ *et al.* (1992). Expression and characterization of human FKBP52, an immunophilin that associates with the 90-kDa heat shock protein and is a component of steroid receptor complexes. *Proc Natl Acad Sci U S A* **89:** 10974-10978.

Perola E, Walters WP, Charifson PS (2004). A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **56:** 235-249.

Peschke M, Blades A, Kebarle P (2002). Charged states of proteins. Reactions of doubly protonated alkyldiamines with NH(3): solvation or deprotonation. Extension of two proton cases to multiply protonated globular proteins observed in the gas phase. *J Am Chem Soc* **124:** 11519-11530.

Peschke M, Verkerk UH, Kebarle P (2004a). Features of the ESI mechanism that affect the observation of multiply charged noncovalent protein complexes and the determination of the association constant by the titration method. *J Am Soc Mass Spectrom* **15:** 1424-1434.

Peschke M, Verkerk UH, Kebarle P (2004b). Prediction of the charge states of folded proteins in electrospray ionization. *Eur J Mass Spectrom (Chichester, Eng)* **10:** 993-1002.

Poulter MO, Payne KB, Steiner JP (2004). Neuroimmunophilins: a novel drug therapy for the reversal of neurodegenerative disease? *Neuroscience* **128:** 1-6.

Powell HR (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* **55:** 1690-1695.

Price NC, Dwek RA, Ratcliffe RGWMR. (2007). Physical Chemistry for Biochemists.

Rarey M, Kramer B, Lengauer T, Klebe G (1996). A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **261:** 470-489.

Rayleigh L (1882) *Philosophical magazine* **14:** 184.

Reynolds CH, Tounge BA, Bembenek SD (2008). Ligand binding efficiency: trends, physical basis, and implications. *J Med Chem* **51:** 2432-2438.

Robinson CV, Chung EW, Kragelund BB, Knudsen J, Aplin RT, Poulsen FM *et al.* (1996). Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein-CoA Ligand Binding and Assembly. *J Am Chem Soc* **118:** 8646-8653.

Rogniaux H, Van DA, Barth P, Biellmann JF, Barbanton J, van ZM *et al.* (1999). Binding of aldose reductase inhibitors: correlation of crystallographic and mass spectrometric studies. *J Am Soc Mass Spectrom* **10:** 635-647.

Rohrig CH, Loch C, Guan JY, Siegal G, Overhand M (2007). Fragment-Based Synthesis and SAR of Modified FKBP Ligands: Influence of Different Linking on Binding Affinity. *ChemMedChem* **2:** 1054-1070.

Rosen MK, Standaert RF, Galat A, Nakatsuka M, Schreiber SL (1990). Inhibition of FKBP rotamase activity by immunosuppressant FK506: twisted amide surrogate. *Science* **248:** 863-866.

Rulten SL, Kinloch RA, Tateossian H, Robinson C, Gettins L, Kay JE (2006). The human FK506-binding proteins: characterization of human FKBP19. *Mamm Genome* **17:** 322-331.

Sanderson DM, Earnshaw CG (1991). Computer prediction of possible toxic action from chemical structure; the DEREK system. *Hum Exp Toxicol* **10:** 261-273.

Scaronolc k., Stockmayer W.H. (1973). Kinetics of diffusion-controlled reaction between chemically asymmetric molecules. II. Approximate steady-state solution. *Int J Chem Kinet* **5:** 733-752.

Schnecke V, Bostrom J (2006). Computational chemistry-driven decision making in lead generation. *Drug Discov Today* **11:** 43-50.

Scholz C, Zarnt T, Kern G, Lang K, Burtscher H, Fischer G *et al.* (1996). Autocatalytic folding of the folding catalyst FKBP12. *J Biol Chem* **271:** 12703-12707.

Schuffenhauer A, Brown N, Selzer P, Ertl P, Jacoby E (2006). Relationships between Molecular Complexity, Biological Activity, and Structural Diversity. *J Chem Inf Model* **46:** 525-535.

Seiler KP, George GA, Happ MP, Bodycombe NE, Carrinski HA, Norton S *et al.* (2008). ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res* **36:** D351-D359.

Shave S, Taylor P, Walkinshaw M (2008) *IBM, Journal of Research and Development:Applications of massively parallel systems* **'in press'**.

Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274:** 1531-1534.

Sich C, Improta S, Cowley DJ, Guenet C, Merly JP, Teufel M *et al.* (2000). Solution structure of a neurotrophic ligand bound to FKBP12 and its effects on protein dynamics. *Eur J Biochem* **267:** 5342-5355.

Siekierka JJ, Hung SH, Poe M, Lin CS, Sigal NH (1989). A cytosolic binding protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin. *Nature* **341:** 755-757.

Sinars CR, Cheung-Flynn J, Rimerman RA, Scammell JG, Smith DF, Clardy J (2003). Structure of the large FK506-binding protein FKBP51, an Hsp90-binding protein and a component of steroid receptor complexes. *Proc Natl Acad Sci U S A* **100:** 868-873.

Snowden M, Green DV (2008). The impact of diversity-based, high-throughput screening on drug discovery: "chance favours the prepared mind". *Curr Opin Drug Discov Devel* **11:** 553-558.

Somarelli JA, Lee SY, Skolnick J, Herrera RJ (2008). Structure-based classification of 45 FK506-binding proteins. *Proteins* **72:** 197-208.

Spencer SD, Abdul O, Schulingkamp RJ, Raffa RB (2002). Toward the design of ribonuclease (RNase) inhibitors: ion effects on the thermodynamics of binding of 2'-CMP to RNase A. *J Pharmacol Exp Ther* **301:** 925-929.

Spolar RS, Ha JH, Record MT, Jr. (1989). Hydrophobic effect in protein folding and other noncovalent processes involving proteins. *Proc Natl Acad Sci U S A* **86:** 8382-8385.

Stahl M, Mauser H, Tsui M, Taylor NR (2005). A robust clustering method for chemical structures. *J Med Chem* **48:** 4358-4366.

Stebbins JL, Zhang Z, Chen J, Wu B, Emdadi A, Williams ME *et al.* (2007). Nuclear magnetic resonance fragment-based identification of novel FKBP12 inhibitors. *J Med Chem* **50:** 6607-6617.

Steinberg TH, Jones LJ, Haugland RP, Singer VL (1996). SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Anal Biochem* **239:** 223-237.

Steiner JP, Dawson TM, Fotuhi M, Glatt CE, Snowman AM, Cohen N *et al.* (1992). High brain densities of the immunophilin FKBP colocalized with calcineurin. *Nature* **358:** 584-587.

Steiner JP, Galey D, Haughey NJ, Asch D, Nath A (2007). Neuroprotective and antiretroviral effects of the immunophilin ligand GPI 1046. *J Neuroimmune Pharmacol* **2:** 49-57.

Stierand K, Maass PC, Rarey M (2006). Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. *Bioinformatics* **22:** 1710-1716.

Sturtevant JM (1977). Heat capacity and entropy changes in processes involving proteins. *Proc Natl Acad Sci U S A* **74:** 2236-2240.

Sun F, Li P, Ding Y, Wang L, Bartlam M, Shu C *et al.* (2003). Design and structure-based study of new potential FKBP12 inhibitors. *Biophys J* **85:** 3194-3201.

Sun J, Kitova EN, Sun N, Klassen JS (2007). Method for identifying nonspecific protein-protein interactions in nanoelectrospray ionization mass spectrometry. *Anal Chem* **79:** 8301-8311.

Taylor P, Blackburn E, Sheng YG, Harding S, Hsin KY, Kan D *et al.* (2008). Ligand discovery and virtual screening using the program LIDAEUS. *Br J Pharmacol* **153 Suppl 1:** S55-S67.

Teague SJ, Davis AM, Leeson PD, Oprea T (1999). The Design of Leadlike Combinatorial Libraries. *Angew Chem Int Ed Engl* **38:** 3743-3748.

Tellinghuisen J (2003). A study of statistical error in isothermal titration calorimetry. *Anal Biochem* **321:** 79-88.

Tetko IV, Tanchuk VY (2002). Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *J Chem Inf Comput Sci* **42:** 1136-1145.

Thomas MP, McInnes C, Fischer PM (2006). Protein structures in virtual screening: a case study with CDK2. *J Med Chem* **49:** 92-104.

Tjernberg A, Carno S, Oliv F, Benkestock K, Edlund PO, Griffiths WJ *et al.* (2004). Determination of dissociation constants for protein-ligand complexes by electrospray ionization mass spectrometry. *Anal Chem* **76:** 4325-4331.

Todeschini R., Consonni V. (2005). Handbook of Molecular Descriptors. Wiley-VCH, UK.

Touboul D, Jecklin MC, Zenobi R (2008a). Investigation of deprotonation reactions on globular and denatured proteins at atmospheric pressure by ESSI-MS. *J Am Soc Mass Spectrom* **19:** 455-466.

Touboul D, Jecklin MC, Zenobi R (2008b). Ion internal energy distributions validate the charge residue model for small molecule ion formation by spray methods. *Rapid Commun Mass Spectrom* **22:** 1062-1068.

Valentine H, Chen Y, Guo H, McCormick J, Wu Y, Sezen SF *et al.* (2007). Neuroimmunophilin ligands protect cavernous nerves after crush injury in the rat: new experimental paradigms. *Eur Urol* **51:** 1724-1731.

Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1991). Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science* **252:** 839-842.

Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229:** 105-124.

Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD *et al.* (2004). Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J Chem Inf Comput Sci* **44:** 793-806.

Walensky LD, Gascard P, Fields ME, Blackshaw S, Conboy JG, Mohandas N *et al.* (1998). The 13-kD FK506 binding protein, FKBP13, interacts with a novel homologue of the erythrocyte membrane cytoskeletal protein 4.1. *J Cell Biol* **141:** 143-153.

Wang HQ, Nakaya Y, Du Z, Yamane T, Shirane M, Kudo T *et al.* (2005). Interaction of presenilins with FKBP38 promotes apoptosis by reducing mitochondrial Bcl-2. *Hum Mol Genet* **14:** 1889-1902.

Wang XJ, Etzkorn FA (2006). Peptidyl-prolyl isomerase inhibitors. *Biopolymers* **84:** 125-146.

Waszkowycz B (2008). Towards improving compound selection in structure-based virtual screening. *Drug Discov Today* **13:** 219-226.

Wear MA, Kan D, Rabu A, Walkinshaw MD (2007a). Experimental determination of van der waals energies in a biological system. *Angew Chem Int Ed Engl* **46:** 6453-6456.

Wear MA, Patterson A, Malone K, Dunsmore C, Turner NJ, Walkinshaw MD (2005). A surface plasmon resonance-based assay for small molecule inhibitors of human cyclophilin A. *Anal Biochem* **345:** 214-226.

Wear MA, Patterson A, Walkinshaw MD (2007b). A kinetically trapped intermediate of FK506 binding protein forms in vitro: chaperone machinery dominates protein folding in vivo. *Protein Expr Purif* **51:** 80-95.

Wear MA, Walkinshaw MD (2006). Thermodynamics of the cyclophilin-A/cyclosporin-A interaction: a direct comparison of parameters determined by surface plasmon resonance using Biacore T100 and isothermal titration calorimetry. *Anal Biochem* **359:** 285-287.

Wear MA, Walkinshaw MD (2007). Determination of the rate constants for the FK506 binding protein/rapamycin interaction using surface plasmon resonance: an alternative sensor surface for Ni2+-nitrilotriacetic acid immobilization of His-tagged proteins. *Anal Biochem* **371:** 250-252.

Wedemeyer WJ, Welker E, Scheraga HA (2002). Proline cis-trans isomerization and protein folding. *Biochemistry (Mosc)* **41:** 14637-14644.

Wei L, Wu Y, Wilkinson DE, Chen Y, Soni R, Scott C *et al.* (2002). Solid-phase synthesis of FKBP12 inhibitors: N-sulfonyl and N-carbamoylprolyl/pipecolyl amides. *Bioorg Med Chem Lett* **12:** 1429-1433.

Weininger D (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* **28:** 31-36.

Willett P (2006). Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today* **11:** 1046-1053.

Wilson KP, Yamashita MM, Sintchak MD, Rotstein SH, Murcko MA, Boger J *et al.* (1995). Comparative X-ray structures of the major binding protein for the immunosuppressant FK506 (tacrolimus) in unliganded form and in complex with FK506 and rapamycin. *Acta Crystallogr D Biol Crystallogr* **51:** 511-521.

Woods AS, Ferre S (2005). Amazing stability of the arginine-phosphate electrostatic interaction. *J Proteome Res* **4:** 1397-1402.

Wu B, Li P, Liu Y, Lou Z, Ding Y, Shu C *et al.* (2004). 3D structure of human FK506-binding protein 52: implications for the assembly of the glucocorticoid receptor/Hsp90/immunophilin heterocomplex. *Proc Natl Acad Sci U S A* **101:** 8348-8353.

Wu SY, McNae I, Kontopidis G, McClue SJ, McInnes C, Stewart KJ *et al.* (2003). Discovery of a novel family of CDK inhibitors with the program LIDAEUS: structural basis for ligand-induced disordering of the activation loop. *Structure* **11:** 399-410.

Wyatt PJ (1993). Light scattering and the absolute characterization of macromolecules. *Anal Chim Acta* **272:** 1-40.

Yang WM, Yao YL, Seto E (2001). The FK506-binding protein 25 functionally associates with histone deacetylases and with transcription factor YY1. *EMBO J* **20:** 4814-4825.

Yang Y, Moir E, Kontopidis G, Taylor P, Wear MA, Malone K *et al.* (2007). Structure-based discovery of a family of synthetic cyclophilin inhibitors showing a cyclosporin-A phenotype in Caenorhabditis elegans. *Biochem Biophys Res Commun* **363:** 1013-1019.

Yin S, Xie Y, Loo JA (2008). Mass spectrometry of protein-ligand complexes: enhanced gas-phase stability of ribonuclease-nucleotide complexes. *J Am Soc Mass Spectrom* **19:** 1199-1208.

Zhao L, Huang W, Liu H, Wang L, Zhong W, Xiao J *et al.* (2006). FK506-binding protein ligands: structure-based design, synthesis, and neurotrophic/neuroprotective properties of substituted 5,5-dimethyl-2-(4-thiazolidine)carboxylates. *J Med Chem* **49:** 4059-4071.

Zhao Y, Ke H (1996). Mechanistic implication of crystal structures of the cyclophilin-dipeptide complexes. *Biochemistry (Mosc)* **35:** 7362-7368.